Essays in Public and Behavioural Economics

by

Daniel Díez Alonso

Thesis

Submitted to the University of Warwick

for the Degree of

Doctor of Philosophy in Economics

Department of Economics

September 2021

Contents

List of Figures

List of Tables

Acknowledgements

I would like to take this opportunity to express my gratitude to everyone who has helped me throughout my years as a doctoral student.

First, I would like to thank my supervisors Prof. Carlo Perroni, Prof. Lucie Gadenne, and Prof. Miguel Almunia, for their continuous support and guidance over the past four years.

I am also extremely grateful to the faculty and doctoral students of the Economics Department at the University of Warwick I had the chance to interact with throughout this journey. In particular, I would like to thank Kirill Pogorelskiy for his feedback on experimental design and the group of PhD students doing research in Public Economics for our fruitful discussions during our lockdown online coffee breaks. I am also very thankful to my friends who offered their time to discuss my research ideas and provide invaluable feedback. Special mention to Jon, Roberto, Andrei, Neha, and Kathi.

Finally, I thank my extended family: my parents, grandparents, brother, flatmates, partners, and close friends who have suffered with me and provided a much-needed emotional support for the last six years. I am fortunate to have you in my life.

Declaration

This thesis is submitted to the University of Warwick in accordance with the requirements of the degree of Doctor of Philosophy in Economics. I declare that it has not been submitted for a degree at another university. I am the sole author of the three chapters included in this thesis.

September 2021

Abstract

This thesis centres around two main topics of public and behavioural economics: the sources of bias in perceived income ranks and the determinants of pass-through of soda taxes.

Chapter 1 explores the role that income taxes play in shaping taxpayers' perceptions of income distributions. Contrary to the predictions of classical models, low-income people tend to display lower support for redistribution due to a biased perception of society's income distribution. Can income taxes themselves amplify or reduce that bias? I present a theoretical framework that maps the information from the tax schedule onto informative signals that taxpayers use to infer their perceived position in the income distribution. Running probabilistic regression analysis on data from the General Social Survey, I find evidence supporting that changes in the federal income tax system influenced the perceptions of income ranks observed in the USA in the last decades. To identify causality in a controlled environment, I then test the main predictions of the theoretical model by randomising tax systems in an online experiment conducted on Amazon Mechanical Turk with American workers. The results of a large pilot identify statistically significant differences between individuals facing a proportional tax system with a unique average tax rate and those facing a progressive tax system with increasing marginal tax rates. Compared to no tax information (control), facing the progressive tax system used in the experiment induced a 12% higher perceived average income level and a 25% lower perceived probability of being above the average income level among low-income individuals. In contrast, the proportional system did not generate significant differences. These findings encourage further research to identify the exact elements in a tax schedule that generate a bias that can affect support for redistributive policies.

Chapter 2 evaluates the pass-through of sugar taxes on soft drinks in online markets. The UK introduced the Soft Drinks Industry Levy (SDIL), a sort of sugar tax levied on producers, in April 2018. Using web-scraped daily prices from Amazon for non-alcoholic beverages sold in the UK, I explore the pass-through of such tax to online market prices of soft drinks. In addition to the traditional measure of average pass-through, I estimate the impact of the tax on prices of direct (untaxed) substitutes, and I compare categories and package sizes facing different demand elasticity and levels of market concentration. My results confirm the predictions of a general model of tax pass-through in a market with imperfect

competition and product differentiation. The UK sugar drinks affected by the SDIL experienced full pass-through of the tax to consumer prices on average, with over-shifting on middle-sized packages (400ml-999ml). The impact of the tax was larger on prices of cola drinks, the category with lower demand elasticity and higher market concentration. At the same time, prices of sugar-free alternatives also increased by nearly 40% of the tax value, despite being exempted from it.

Chapter 3 investigates the impact of psychological pricing on the pass-through of soda taxes. Following the path of a growing number of countries and the recommendations of the WHO, Spain increased taxes on sweetened soft drinks on 1st January 2021. Nevertheless, the evidence on pass-through of such corrective taxes is mixed, with estimates ranging from 30% to above 100%. Furthermore, psychological pricing is widespread in grocery stores, with over 60 per cent of product prices with *odd endings* (e.g. .49 or .99) in some countries, and has been linked to price rigidity in other settings. Using web-scraped prices data from three major supermarket chains in Spain, I estimate that psychological prices are a significant determinant of tax pass-through. Looking at the last digit of prices, I find that the VAT tax increase on sweetened drinks overcame the price rigidity of zero and nine-ending prices observed in control items. Indeed, products initially priced at round endings (double zero) over-shifted the tax on consumer prices. On the other hand, items initially priced around the half of a Euro (with the cents digits between 40 and 58) experienced partial pass-through levels below 75 per cent. This implies that corrective taxes may result in more significant price increases in markets with a higher frequency of round endings, even if those endings usually increase the price-rigidity of products facing smaller cost shocks. It also confirms that a tax that accounts for ten per cent of product prices is enough to raise the prices of unhealthy drinks significantly, overcoming the rigidity effect of psychological price endings.

1 Taxpayer bias in perceived income distributions

1.1 Introduction

Standard models of preferences for redistribution based on the median voter theorem predict that societies with average income above the median should implement redistributive policies since a majority of the population could benefit from them (Meltzer and Richard, 1981). Given the levels of inequality observed in most western democracies, however, empirical evidence shows lower support for redistribution than expected, creating what the literature has called the *inequality-redistribution puzzle* (Benabou, 1996; Kenworthy and McCall, 2007).

From a theoretical perspective, the fact that changes in income inequality in a society rarely translate into changes in preferences for redistribution remains a puzzle. Part of the literature rationalised this apparent inconsistency by introducing other variables influencing taxpayers' preferences, such as the *prospects of upward mobility* (POUM) hypothesis (Bénabou and Ok, 1998). Nevertheless, those models still rely on individuals effectively inferring the income distribution of their society and their position within it, which is most often not the case. Several studies[1] have consistently found that taxpayers actually have biased perceptions of the income distribution. Moreover, their levels of preferred redistribution seem to respond to those biased perceptions rather than their true position, which can be corrected by giving them accurate information on their actual position or the real level of inequality (Cruces et al., 2013; Kuziemko et al., 2015; Karadja et al., 2017; Hvidberg et al., 2020).

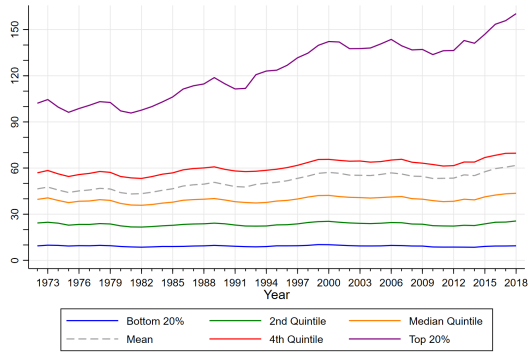
The United States case illustrates this problem very well. Income inequality in the USA has been steadily increasing over the past 50 years and, while citizens seem to be aware of such a trend, their perceived distance from the average family income has barely changed (Figure 1.1). Similarly, support for redistribution has not increased, especially among families with the lowest levels of income, who would benefit the most from redistributive policies (Figure 1.2). Meanwhile, income taxes in the USA have been significantly reduced over the same period, especially for the rich. Have changes in preferences led to such changes in policy, or may the policies themselves shape preferences?

As Gimpelson and Treisman (2018) consistently estimate for a wide range of coun-

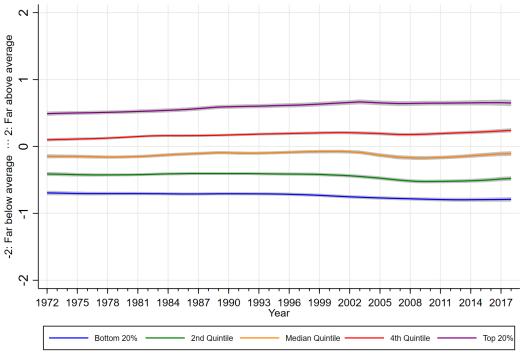
[1] Kenworthy and McCall (2007); Cruces et al. (2013); Hauser and Norton (2017); Fernández-Albertos and Kuo (2018); Hvidberg et al. (2020).

tries using ISSP[2] data, people's support for redistribution seems to be driven by their perceived income distribution rather than the true one. Therefore, understanding what shapes those perceptions is a key political question. So far, the focus has been on analysing the role of reference groups (Cruces et al., 2013; Hvidberg et al., 2020), treating tax systems as the mere result of taxpayers' perceptions and preferences. However, what if tax schedules generated reference points that influenced income perception biases? This paper adds a new perspective to the discussion: if taxpayers infer distributional information from tax policies, the policies themselves may generate a bias that affects their public support.

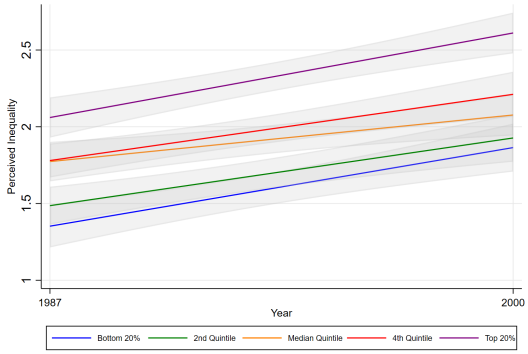
Figure 1.1: Evolution of household incomes, perceived position, and perceived inequality



(a) Family Income Distribution



(b) Perceived distance from the average family income

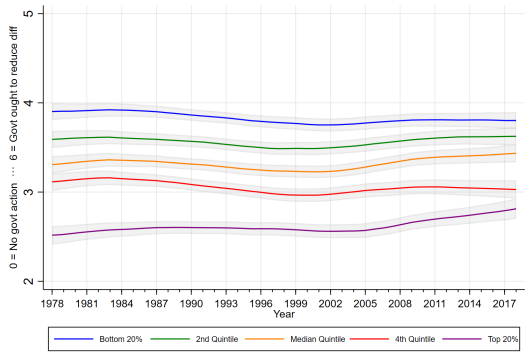


(c) Perceived pay inequality

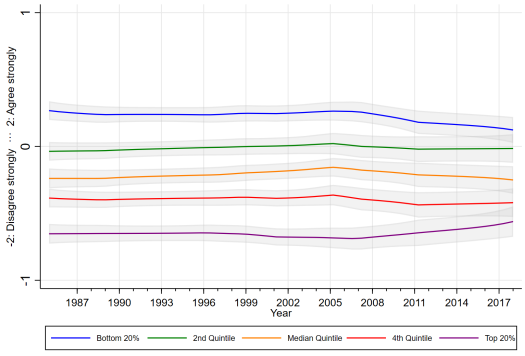
Source: Elaborated by the author using General Social Survey (GSS) and Census data.
Notes: Figure 1.1a shows the evolution of the average income per quintile and depicts the increase in inequality throughout the last 50 years. Figure 1.1b shows how families in each income quintile perceive their income compares to the average. Figure 1.1c shows the change in perceived pay inequality ($\ln \frac{perceivedexecutivepay}{perceivedskilledworkerpay}$) between 1987 and 2000, by income quintile.

[2]The International Social Survey Programme (ISSP) conducts annual surveys covering countries from North and South America, Africa, Europe, Asia and Oceania, with a module on social inequality.

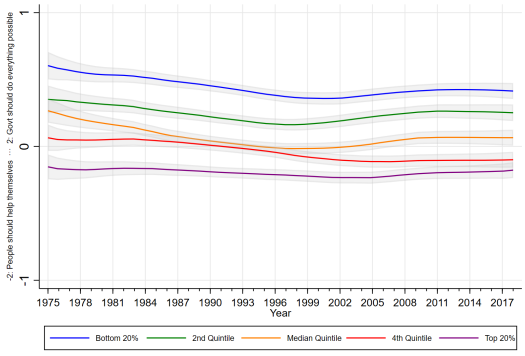
Figure 1.2: Evolution of preferences for redistribution, by income quintiles



(a) Should the government act to reduce differences between rich and poor?



(b) It is the government's responsibility to reduce income differences between rich and poor.



(c) Should the government improve the standard of living of poor people?

Source: Elaborated by the author using GSS and Census data.

Notes: The three graphs in this figure summarise the evolution of the answers to three questions included in the General Social Survey (GSS) through the last four to five decades. The answer to question (a) was a scale between zero (*No governmental action*) and six (*Government ought to reduce differences*). Question (b) was answered on a five-point scale, between minus two (*Disagree strongly*) and plus two (*Agree strongly*). The answer to question (c) was also a five-point scale, in this case recording positions between *People should help themselves* (minus two) and *Government should do everything possible* (plus two).

Some behavioural economics papers have built on classical models to explain some of the most common biases, such as the representativeness heuristic when inferring population distributions from a reduced sample (Cruces et al., 2013). However, empirical evidence from survey data in the USA suggests that the predictions of those models do not explain the behaviour at the bottom tail of the income distribution. The representativeness heuristic in this context implies that individuals believe their self-selected reference group is more representative of the whole population than it really is, leading them to feel closer to the average than they actually are. In a model with such base-rate neglect, the probability of an individual perceiving herself as earning the average income should decrease as

income approaches the boundaries (zero at the bottom and the highest possible income at the top) since the probability that other incomes in their reference group extend beyond those boundaries is zero. Nevertheless, I observe that the left tail of the income distribution does not behave this way. Specifically, it seems that all respondents with family incomes below a level that coincides with the income tax personal allowance (i.e. all levels of income exempt from paying federal income tax) perceive they are at the same distance from the average family income, regardless of their level of earnings. I argue that the tax-free allowance sets a reference value that influences individuals' perceived position.

If this hypothesis is true, when thresholds of the tax bands or their marginal tax rates change, poorer households may perceive themselves closer to the average family income and therefore reduce their support for redistributive policies, allowing the social planner to cut taxes on higher-income groups. This is, in fact, what one observes when analysing the evolution of the Federal Income Tax Law in the USA for the last half of the 20th century: despite rising inequality, the government approved significant tax cuts for top incomes through several reforms of the federal income tax schedule. Simultaneously, poor households' perceived distance from the average family income and their level of support for redistributive policies barely changed. Two questions arise from those facts. First, why did low-income families fail to update their perceived distance from the average income, even when they perceived increasing pay inequalities in their society? And second, why did they not increase their support for redistributive policies?

In this study, I use data from the General Social Survey (GSS) to identify patterns on different measures related to perceived income distributions which the existing models with reference group bias cannot explain. I present a theoretical framework that builds on existing behavioural models and maps information from the income tax schedule onto informative signals that taxpayers use to infer income ranks, what I call the *tax burden heuristic*. I then test the model's predictions in an online experiment using Amazon Mechanical Turk, with participants based in the USA.

The final experiment of the paper finds statistically significant differences between individuals facing a progressive tax system with increasing marginal tax rates (informative) compared to those facing a proportional tax system with a unique flat rate (uninformative) and those in the control group who do not have any information about the tax system. In particular, taxpayers in the progressive tax group believe they are approximately 25% less likely to be above the average income in their reference group than those in the control and the proportional tax

groups. They also estimate the average income to be around 12% higher. This finding is consistent with the hypothesis presented in this paper: setting different contribution levels above specific income thresholds affects how taxpayers infer their position in the income distribution.

This work contributes to the growing literature on (mis)perceptions of inequality. The new theory I present helps to explain the *inequality-redistribution puzzle* and provide a better understanding of the political economics of redistributive preferences, building on the work of Piketty (1995), Alesina and Giuliano (2011) and Cruces et al. (2013). Moreover, the conclusions of this study may also have implications for the literature of economics of happiness, which sustains that happiness is affected by relative income (rank) rather than absolute income levels (Boyce et al., 2010). Additionally, the notion that income tax thresholds constitute reference points that trigger reactions beyond labour supply and tax avoidance decisions is of relevance to the literature of public finance and should be taken into account when deriving optimal income taxes.

The rest of the paper is structured as follows: Section II develops the theoretical framework and presents the concept of *tax burden heuristic*; Section III provides supporting evidence from USA survey data, exploiting tax law changes that generated quasi-experimental variation; Section IV presents the online experiment aimed to test the predictions of the model; Section V concludes.

1.2 Theoretical framework

One of the most consistent biases in perceived income rank is that people tend to feel closer to the average income than they really are. To explain that bias, Cruces et al. (2013) derived a model in which taxpayers suffer from base-rate neglect[3] and therefore take the income distribution of their reference group (to which they are self-selected by similarity) as if it was representative of the overall population. This leads people to believe they are closer to the average income than they truly are, as depicted in Figure 1.3. In the example of the figure, the blue solid line represents a hypothetically true cumulative distribution of incomes. The red dashed line would be the distribution of beliefs regarding income rank if income were bounded at both ends (minimum and maximum income levels), and respondents based their perception on their reference group, to which they are self-selected by similarity. Notice that although the bias increases as incomes depart from the average (marked by the grey vertical line at 0.2), it reduces

[3]The concept of base-rate neglect is such of sampling bias described by Kahneman and Tversky (1972).

again as incomes approach either boundary, since the probability of income levels beyond the boundaries is zero (in terms of the reference group, if you earn the minimum possible income, it cannot be that you know someone who earns less). A similar prediction would apply to a question regarding the position of the household relative to the income of the average family, since the probability of being above (below) the average tends to zero as incomes approach the lower (upper) boundary, for any possible distribution with positive densities along the feasible income range and zero density for income values beyond its boundaries. Hvidberg et al. (2020) argue that the bias at the tails may be larger due to mean-reversion since people at the bottom cannot underestimate their position and people at the top cannot overestimate it. However, that same argument would imply that the dispersion of answers should decrease as it approaches those boundaries[4], which is not what we observe in the data. Perceived position at the lower tail, the area that does not behave as predicted by a model with base-rate neglect, presents larger variation than at any other levels of income, perhaps signalling less accurate information.

A question in the GSS specifically asks: *"Compared with American families in general, would you say your family income is...?" [Far below average - Below average - Average - Above average - Far Above average].* In Figure 1.4a, I recoded the answers to show the proportion of people in each income level (normalised by the average family income of each year) perceiving themselves as below, around, or above the average for all survey years between 1984 and 2018. The grey vertical line represents the average income for the year of each survey. As expected, the proportion of people feeling below (above) the average is larger for lower (higher) incomes. Meanwhile, the proportion of people feeling around the average is the highest around the true average family income and decreases as income levels depart from it. It is relevant to notice that the proportion of respondents perceiving their income above the average remains positive and stable (around 10 per cent) at lower income levels, even if this is not consistent with the predictions of models based on similarity-selected reference groups. Similarly, the proportion of people feeling below the average seems fairly stable (at around 60 per cent) for all people earning any income below a quarter of the average family income. Figure 1.4b represents the distribution of perceived position by true income rank. The blue dots and fitted solid line represent the mean value of answers per income group in each year, while the green dashed line serves as a reference value for a

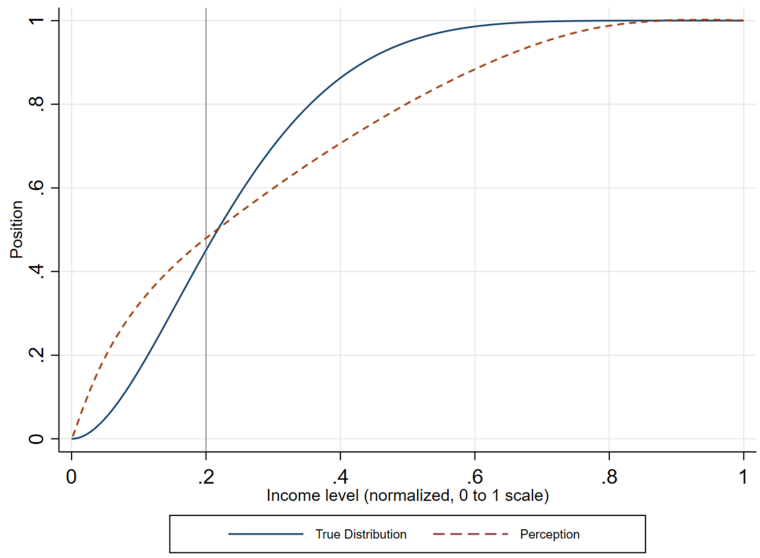
[4]The dispersion could, however, increase for other reasons, hence challenging this argument. Nevertheless, the correlation between income and the perception bias revealed by the GSS data has a clear inflexion point way before the answers are close to the extremes of the scale, rather than flattening smoothly as answers get closer to their lower bound (see Figure 1.4).

calculated measure of true distance from the average family income.

Given the top coding of the income variable in the GSS data, I focus the analysis on the anomaly observed at the left tail of the distribution. The average answer stops decreasing below a certain income level despite incomes being at the lowest end of the distribution, and uncertainty around perceived positions increases. This contradicts the prediction of models with base-rate neglect, which would yield more accurate perceptions at lower income levels, where the probability of being above the average family income tends to zero.

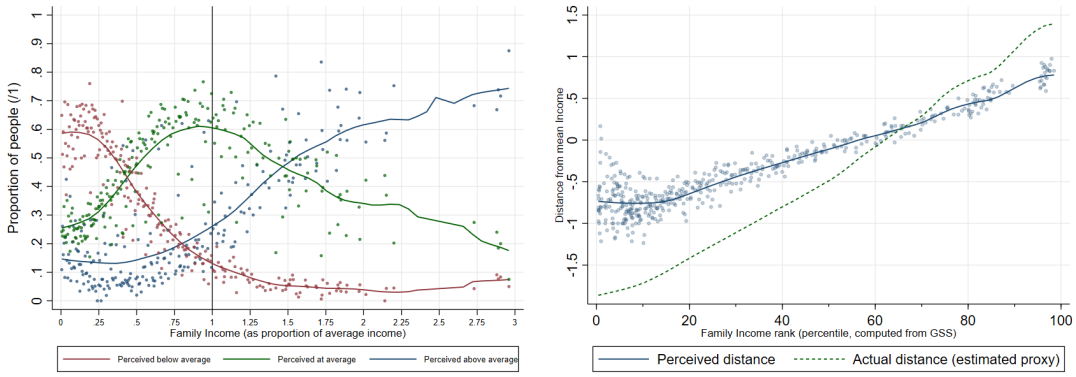
What is the driver of such increased bias at the lower tail? Is there a specific reference value from which people have a fixed perception of their position in the income distribution? I suggest an alternative source of information that may

Figure 1.3: Perceived position under base-rate neglect and self-selected reference group



Notes: In this model, the central assumption is that the reference group generating the base-rate neglect only covers a subsample of the population distribution around the income level of the individual reporting her perceived position. Since incomes in that group are closer to the respondent's income than other incomes in the general population, the respondent underestimates her distance from the average and median incomes. However, for agents at the extremes of the distribution (i.e. with the highest or the lowest possible incomes), the probability of knowing a person with income beyond those boundaries is zero, hence reducing the bias. If *position* refers to the percentile, the lowest possible perceived percentile will depend on the frequency of the minimum income (not necessarily zero). However, if *position* refers to the probability to have an income above the average, as in one of the GSS questions, this should tend to zero as the respondent's income approaches the lower boundary of the income distribution, and it should tend to one as it approaches the other end. The graph reproduces the case for this second meaning of *position* relative to the average income, which is the central measure used in this paper and that of most relevance to redistributive preferences.

Figure 1.4: Perceived position relative to the average family income



(a) By relative income (b) By true income rank

Source: Elaborated by the author using GSS and Census Data.

Notes: These two graphs plot the distribution of answers to the GSS Question 202: *Compared with American families in general, would you say your family income is...*, which has five possible answers: *Far below average; Below average; Average; Above average; Far above average.* The left figure (a) shows the proportion of respondents perceiving their family income below, around, or above the average family income by the true proportion of their family income with respect to the average. Families with the average income are those around value one in the horizontal axis, while those on the left (right) are families with incomes below (above) the average. The blue dots and fitted polynomial (blue line) in the right figure (b) represent the average score of all answers in a given year, with zero denoting respondents who perceive their family income is at the USA average and values one and two (positive and negative) for the different intensities of the above and below average answer options. The horizontal axis in the second graph denotes the true percentile of the respondent's family income. The green dotted line is only plotted for visual reference and is an arbitrary (albeit proportional) measure of distance from the average family income based on the relative size of the family income at each percentile of the distribution compared to the average income. Alternative measures could be computed, leading to different slopes, but they would all cross the true average income level (zero on the vertical axis) at the same point, slightly above the 60th percentile, and cross the blue fitted line afterwards.

be shaping those perceptions: the tax burden.[5] Do taxpayers infer their income position from how much they pay in taxes rather than uniquely from the value of their income? Certainly, in any progressive tax system with increasing marginal tax rates, people earning the least will pay the lower proportion of their income in taxes, while those earning top incomes will be taxed at a higher rate (larger share). The tax burden is exactly the same for all income levels below the tax-free allowance (zero). If they were to infer their position based on how much they pay in taxes rather than how much they earn, all those in the tax-free band would perceive themselves at the same position, despite their differences in income.

To formalise the hypothesis, I present a model that rationalises why taxpayers

[5]Throughout this paper, I refer to the proportion of income paid in taxes as the *tax burden*.

may infer their position in the income distribution from tax schedules and, specifically, their tax burden (average income tax rate). The model's main prediction is that a higher (lower) tax burden will be used as a signal of a higher (lower) level of relative income[6] and a higher (lower) rank in the income distribution.

Assume a country is ruled by a government that has to provide public services for its population, and it needs to fund such public expenditure G through tax revenues. In the simplest case, assume the only tax in place is the income tax, so $\sum_i (\tau_i \cdot y_i)$ represents the total budget, collected from each individual's income y_i at an average tax rate τ_i. If individuals were subject to a non-progressive (purely proportional) system, with income taxed at a flat rate $\tilde{\tau}$, the flat rate would be determined by $\frac{G}{Y}$, where Y represents the total addition of individual incomes in the population, $\sum_i (y_i)$. The total tax revenue collected by the tax authority given a tax schedule S thus can be expressed as $G(Y \mid S) = \tilde{\tau} \cdot Y$.

However, in most countries, the income tax schedule is composed of a progressive system of thresholds and increasing marginal tax rates above each income threshold. Such a system is thus defined by a set of *K+1* thresholds ($z_0 = 0 < z_1 < \cdots < z_K$) and a corresponding set of marginal tax rates ($r_0 < r_1 < \cdots < r_K$) where any marginal tax rate r_k is applied to any earned income between z_k and z_{k+1}.[7] The tax burden of an individual, her proportion of income paid in taxes, is measured by the average tax rate (τ_i) as in Equation 1.1, and the total tax revenue can be expressed in this case as $G(Y \mid S) = \int_0^\infty y\,\tau(y)\,f_y\,dy = \sum_{n=1}^{N} y_n\,\tau(y_n)$, where N is the total number of people subject to the tax system and f_y the number of them earning income y.

$$\tau(y_i) = \tau_i = \frac{\sum_{k=0}^{K}\{r_k \cdot min\,(z_{k+1} - z_k, max\,(0, y_i - z_k))\}}{y_i} \quad \text{where } z_{K+1} = +\infty \tag{1.1}$$

Therefore, in any given income tax system, the range of possible tax burdens has a lower bound at 0 and an upper bound equal to the top marginal tax rate, since $\lim_{y_i \to \infty} \tau_i = r_K$. This also means that the reference rate $\tilde{\tau}$ that would apply if all taxpayers faced the same proportional flat rate must lie within $(0, r_K)$.

Any given tax schedule with progressive tax rates must be constructed so that

[6]The term *relative income* is used in this paper as income level relative to the average level of income in the population.

[7]A tax-free allowance would be represented by $r_0 = 0$, and the initial tax band would then kick in at rate r_1 for any unit of income above z_1. Similarly, all units of income above z_K will be taxed at the maximum marginal tax rate, r_K.

those contributing above the hypothetical proportional share $\tilde{\tau}$ compensate the forgone revenue from those contributing below that proportional share. Calling $\tilde{y}$ the income level such that $\tau(\tilde{y}) = \tilde{\tau}$, Equation 1.2 represents such balance condition:

$$\int_0^{\tilde{y}} (\tilde{\tau} - \tau(y))\ y\, f_y\, dy = \int_{\tilde{y}}^{\infty} (\tau(y) - \tilde{\tau})\ y\, f_y\, dy \tag{1.2}$$

If agent i earns income y_i with a resulting tax burden $\tau(y_i) < \tilde{\tau}$, Equation 1.2 can be re-written as Equation 1.3, which establishes the relation in terms of transfers between incomes below and above that of agent i to ensure a target revenue G.

$$\int_0^{y_i} (\tilde{\tau} - \tau(y))\ y\, f_y\, dy + \int_{y_i}^{\tilde{y}} (\tilde{\tau} - \tau(y))\ y\, f_y\, dy = \int_{\tilde{y}}^{\infty} (\tau(y) - \tilde{\tau})\ y\, f_y\, dy$$

$$\int_0^{y_i} (\tilde{\tau} - \tau(y))\ y\, f_y\, dy = \int_{y_i}^{\infty} (\tau(y) - \tilde{\tau})\ y\, f_y\, dy \tag{1.3}$$

1.2.1 Theoretical predictions

Imagine a society with only three types of agents: those earning a high income (y^H), those with medium income (y^M), and those earning a low income (y^L). The government needs to raise a fixed amount G through income taxes. If it implements a proportional tax rate $\tilde{\tau}$ (no progressivity), individuals cannot relate their tax burden to their position in the income distribution (since all individuals pay the same tax rate, independent of their income).

Alternatively, the government could consider a very simple progressive tax system with two marginal tax rates, r_0 up to income level y^*, and then a higher r_1 for any income exceeding that level. This means the possible range of tax burdens for income earners in this society is defined between r_0 and r_1, with $r_0 < \tilde{\tau} < r_1$.

Recalling the balance condition of progressive tax schedules (Equation 1.2), there must be at least one group with a tax burden below the proportional share $\tilde{\tau}$ and at least another group with a tax burden above it. Since the minimum possible tax rate is r_0, which by definition must be smaller than $\tilde{\tau}$, this implies the following information can be inferred by taxpayer i when realising their tax burden, without knowing which income group they belong to, and even without

knowing $\tilde{\tau}$:[8]

$$P(y_i \in L \mid \tau(y_i) = r_0) > 0\,;\; P(y_i \in L \mid \tau(y_i) = r_1) = 0 \tag{1.4}$$

$$P(y_i \in H \mid \tau(y_i) = r_0) = 0\,;\; P(y_i \in H \mid \tau(y_i) = r_1) > 0 \tag{1.5}$$

In addition, assuming $\tilde{\tau}$ is known to taxpayers, the tax system automatically becomes more informative:

$$P\left(y_i \in H \mid r_0 \leqslant \tau(y_i) < \tilde{\tau}\right) = 0 \Rightarrow P(y_i \in M \cup L \mid r_0 \leqslant \tau(y_i) < \tilde{\tau}) = 1 \tag{1.6}$$

$$P\left(y_i \in L \mid \tilde{\tau} < \tau(y_i) \leqslant r_1\right) = 0 \Rightarrow P(y_i \in M \cup H \mid \tilde{\tau} < \tau(y_i) \leqslant r_1) = 1 \tag{1.7}$$

And incorporating the balance condition (1.2):

$$P(y_i \in M \mid \tau(y_i) < \tilde{\tau}) \propto P\left(\left(\tau(y^H) - \tilde{\tau}\right) y^H f^H > \left(\tilde{\tau} - \tau(y_i)\right) y_i \left(N - f^H\right)\right) \tag{1.8}$$

$$P(y_i \in M \mid \tau(y_i) > \tilde{\tau}) \propto P\left(\left(\tilde{\tau} - \tau(y^L)\right) y^L f^L > \left(\tau(y_i) - \tilde{\tau}\right) y_i \left(N - f^L\right)\right) \tag{1.9}$$

Notice that the probabilities in (1.8) and (1.9) both decrease as the distance between $\tau(y_i)$ and $\tilde{\tau}$ increases, everything else constant. Therefore, a lower tax burden below the proportional share implies a lower probability of belonging to income group M (in favour of group L), and a higher tax burden above the proportional share implies a lower probability of belonging to income group M (in favour of group H), for any prior belief on the distribution of incomes. Combining those conditions, this implies the conditional probability of belonging to income group L weakly increases when $\tau(y_i)$ decreases, while the conditional probability of belonging to group H weakly increases when $\tau(y_i)$ increases: $\frac{\partial P\left(y_i \in y^L \mid \tau(y_i)\right)}{\partial \tau(y_i)} \leq 0$ and $\frac{\partial P\left(y_i \in y^H \mid \tau(y_i)\right)}{\partial \tau(y_i)} \geq 0$, everything else constant.

Such calculations of probabilities are complex and thus have a high cognitive cost. However, the qualitative conclusion of this model can be approximated by a heuristic, which I call the *tax burden heuristic*. This heuristic consists in inferring a higher position in the income distribution from a higher *relative* tax burden (τ_i relative to the maximum rate in the tax system or the proportional share $\tilde{\tau}$, when known). Notice that a *higher position* in this model means a larger share of population income accumulated below the respondent's income, which

[8] In these equations, the expression $P(y_i \in L)$ denotes the probability that income y_i belongs to the group of low income earners.

can imply a higher rank or a higher income relative to the average (larger distance from the average income if above it, and shorter distance if below), or both.

1.3 Empirical evidence in the USA

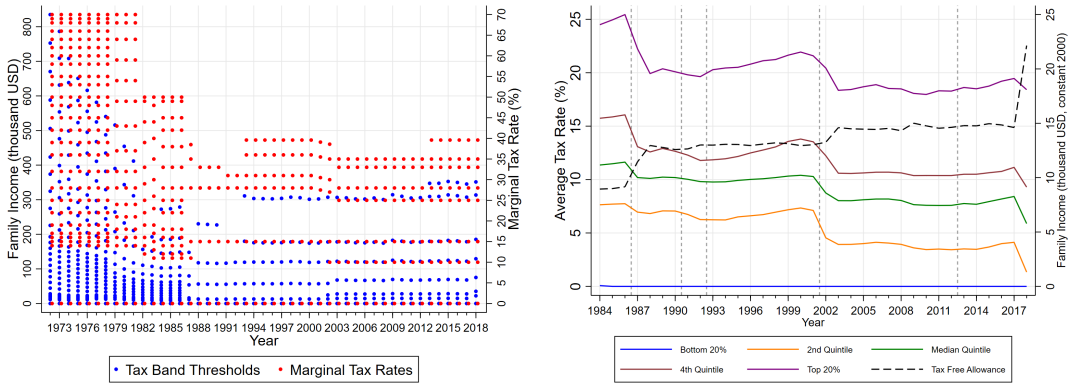
To test the model's predictions, in this section, I show supportive evidence from the USA using household panel data from the General Social Survey (GSS) for years 1984-2018, exploiting the exogenous variation of tax burdens introduced by the Income Tax Law changes within this period. Although the GSS data is available for years as early as 1972, I restrict my analysis to the period from 1984, since Personal Income Tax thresholds were only indexed to inflation from that time. While this survey has its limitations regarding the accuracy of income measures (the public dataset groups households by income ranges rather than disclosing the exact answer from each household, which results in top coding), it is the only publicly available longitudinal dataset including income data, perceived position relative to the average family income, and preferences for redistribution. It also includes the more ambiguous question on "position in the social ladder" from the ISSP module.

1.3.1 Historical context and main reforms

The Federal Personal Income Tax Law in the United States has experienced massive changes during the last 50 years. As depicted in Figure 1.5a, tax thresholds were only indexed to inflation after 1984. This resulted in progressive changes in the real value of those thresholds during the first ten years represented in the graph, although these were most probably not as salient for taxpayers as those resulting from changes in the tax law. All income values are counted in real terms, as constant US dollars with base year 2000. Each blue dot represents the threshold for each additional tax band, and the red dots depict their corresponding marginal tax rates. One can observe a very relevant reduction in the number of tax bands (blue dots) for the first third of the series, followed by a slow progressive increase in subsequent years. Due to the scale of the graph, one cannot notice the changes on the bottom band, the one with zero marginal tax rate (tax-free allowance), but this is represented in the graph on the right, Figure 1.5b. The dashed black line shows the evolution of the tax-free allowance (right vertical axis), and the same chart depicts the evolution of tax burdens for the average family income level within each income quintile (left vertical axis). Changes in the tax-free allowance mechanically have a more significant impact on the burden of lower incomes. However, one can observe how, in most periods

where the tax-free allowance was increased, it was actually the burden of the highest income quintile that was most reduced, implying that those reforms also included significant tax cuts for higher incomes.

Figure 1.5: Evolution of Federal Income Tax for married couples in the USA



(a) Summary of band thresholds and tax rates

(b) Evolution of tax burdens, by income quintiles

Source: Elaborated by the author using Census Data.
Notes: In the left graph (a), the blue dots represent the number of bands set by the Federal Income Tax Law in each year. Their position on the (left) vertical axis represents the threshold for each band. The vertical distribution of these blue dots gives an idea of the piece-linearity of the tax schedule. Meanwhile, the red dots represent the marginal tax rate assigned to each corresponding income band. The chart on the right (b) summarises how tax burdens, measured as the proportion of gross income paid in taxes, changed over time for the average income of each income quintile. The dashed black line (right vertical axis) was added to complement the information included in Figure 1.5a, since it would correspond to the lowest blue dot in every year, indistinguishable from the other bands in most years. The vertical dashed grey lines highlight the years with major changes to the Federal Income Tax Law.

The first relevant change in the income tax during 1972-2018 was introduced by the Tax Reduction and Simplification Act of 1977 and the Revenue Act of 1978, which consisted in removing ten of the intermediate tax thresholds with the main purpose of simplifying the tax schedule. Tax rates were kept constant for the remaining bands, and thus the subtraction of intermediate thresholds effectively resulted in a reduction of taxes on most income levels, especially the highest ones.

The second relevant reform arrived with the Tax Equity and Fiscal Responsibility Act of 1982, which removed the three top tax bands, resulting in a huge drop of the top threshold and corresponding marginal tax rate. That threshold was raised again progressively in the following two years, with the addition of two new bands. The Economic Recovery Tax Act of 1981 also set that all tax thresholds would be adjusted for inflation after 1984, except for the first year after a new law directly affected the value of a tax band.

The most relevant change in the history of income taxes in the USA was implemented in 1987 after the US Congress passed the Tax Reform Act of 1986. The number of tax bands was reduced from 15 to only 6, and tax rates dropped significantly.

A few years later, the Omnibus Budget Reconciliation Act of 1990 started raising the top marginal tax rate, but it was not until 1993 that two additional tax bands were added at the top of the scale by the Omnibus Budget Reconciliation Act of 1993. This reform was followed by the most stable period regarding the Income Tax Law until the Economic Growth and Tax Relief Reconciliation Act of 2001 introduced new tax reductions.

The last big changes of the Income Tax Law took place in 2013, with the introduction of an additional band at the top of the scale, and in 2018, with a substantial increase of the tax-free allowance.

1.3.2 Description of the survey data

The General Social Survey (GSS)has been running in the USA in a panel format since 1972, either annually or biennially, and is publicly available on the website of the NORC (University of Chicago).[9] It is one of the few large panel surveys that has consistently recorded some measure of income rank perception, together with the usual socio-demographic characteristics of respondents. I use this data to find correlations that support the model presented in the previous section to motivate an online experiment that allows testing the model's predictions with identified causal interpretation. I present the results for the period 1984 to 2018, including 22 rounds and comprising 4510 households, of which more than half were interviewed in at least 11 different rounds.

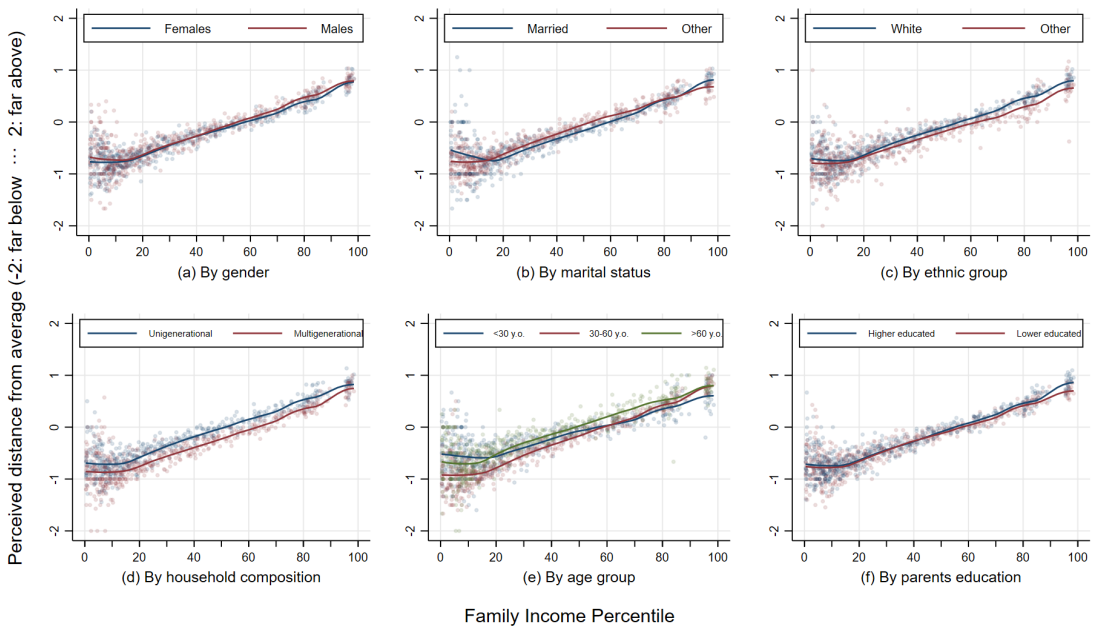
The main dependent variable of interest is the answer to Question 202: *"Compared with American families in general, would you say your family income is...? [Far below average; Below average; Average; Above average; Far above average]"*. In addition, in the years where questions from the International Social Survey Programme (ISSP) were incorporated into the GSS, a similar (but less specific) question is also available. This question has been broadly used for cross-country comparisons in the literature of perceived income rank and preferences for redistribution[10]: *"In our society there are groups which tend to be towards the top and those that are towards the bottom. Here we have a scale that runs from top to*

[9] *http://www.gss.norc.org/* [last accessed on 14/09/2018].
[10] Alesina and Giuliano (2011) and Evans and Kelley (2004), among others.

bottom. Where would you put yourself on this scale? [1=Top ····· 10=Bottom]". I run the analysis on both questions, which I refer to as Q1 and Q2, respectively, through the rest of this section.

As mentioned in the introduction, and as consistently identified in the literature, people below the mean tend to overestimate their position while the reverse happens to people above the mean, and both groups underestimate their distance from the average or median income (Cruces et al., 2013; Hvidberg et al., 2020). More surprisingly, while perceived position (distance from the mean) seems to be correlated with income for most of the range, there is a threshold at the lower (left) tail of the distribution from which perceived position does not decrease with lower income levels. Specifically, the correlation between income and perceived position within the bottom quintile of the income distribution disappears, with the average perceived position being constant for households below the 20th income percentile. This anomaly, incompatible with predictions based on reference groups and base-rate neglect, is consistent across different socio-demographic dimensions (Figure 1.6).

Figure 1.6: Distribution of perceived position, by socio-demographic characteristics



Source: Elaborated by the author using data from the GSS, years 1984 to 2018.
Notes: For each socio-demographic variable, a dot represents the average answer of households within the same percentile in a given year. The solid lines are fitted polynomials (Epanechnikov kernels) superposed for visual clarity.

The ISSP modules also include a question on preferences for redistribution, which I call Q3: *Some people think that the government in Washington ought to reduce*

the income differences between the rich and the poor, perhaps by raising the taxes of wealthy families or by giving income assistance to the poor. Others think that the government should not concern itself with reducing this income difference between the rich and the poor. Think of a score of 1 as meaning that the government ought to reduce the income differences between rich and poor, and a score of 7 meaning that the government should not concern itself with reducing income differences. What score between 1 and 7 comes closest to the way you feel?. In Figure 1.7, I show how answers to this question highly correlate with the inverse of the tax burden.

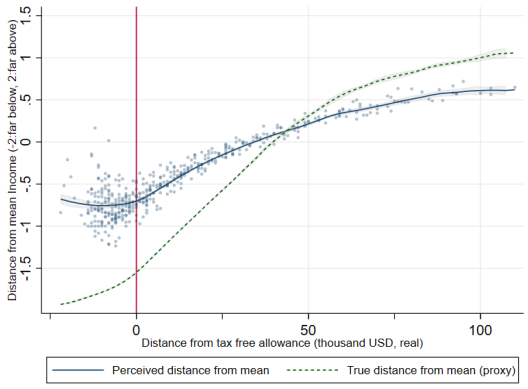
The survey does not include any question on how much taxes the respondent is paying, but I use the yearly family income variable to approximate the corresponding income tax in a given year. Although this measure is not exact due to the lack of information on tax deductions, it is expected to be highly correlated with the real unobserved measure. In the main analysis, I normalise it by the maximum marginal tax rate in that year (potential maximum tax burden as income tends to infinity). This way, I can capture the idea of the relative tax burden, as described in the theoretical model.

The correlation between the answers to each of those three questions and the income level can be observed in Figure 1.7. The income measure on the horizontal axis has been rescaled for every year so that value zero corresponds to the tax-free allowance threshold (red vertical line). One can see that the distribution of answers to each of the three questions (blue) has an inflexion point exactly around the tax-free allowance, which is only observed in the distribution of the tax burden rather than on the actual distributions of percentile and distance from the average income.

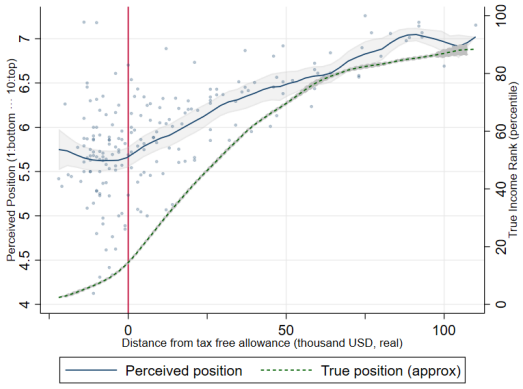
1.3.3 Empirical analysis

Exploiting the exogenous shocks to tax burdens introduced by the changes in the income tax law, I explore the explanatory power of such a source of information to taxpayers. I regress the perceived position (Q1) on the relative tax burden (average tax rate on the respondent's household income, $\tau(y_{it})$, relative to the maximum possible tax rate r_{K_t}) and control for family income (y_{it}), a recent change in financial situation ($FinChange_{it}$), socio-demographic variables, and year (t) fixed effects. Given that the dependent variable is an ordered scale, I estimate the coefficients using an ordered probability model (OProbit), both as cross-sections and as a panel with individual fixed effects (α_i). The sample used

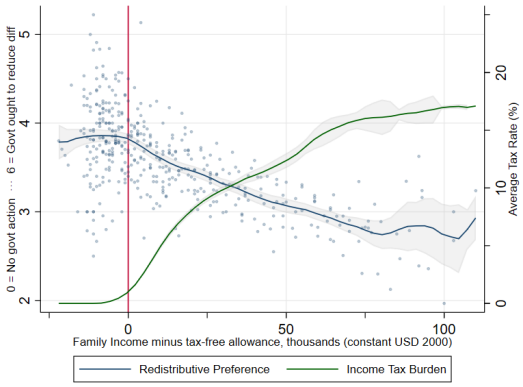
Figure 1.7: Distribution of answers to GSS questions, by distance to personal tax allowance



(a) Q1: Perceived distance from average family income



(b) Q2: Perceived position in society



(c) Q3:Preference for redistributive policies

Source: Created by the author using GSS and Census data for years 1984-2018.

Notes: The blue dots and fitted polynomial in each of the three graphs represent the average answer to three questions of the GSS detailed in previous paragraphs. In these figures, the income of each respondent has been deducted their corresponding personal allowance according to the Income Tax Law in place the year of each survey round. Therefore, the horizontal axis represents the distance between the respondent's household income and the tax-free allowance, adjusted for inflation to 2020 US dollars. The green dashed line in graph (a) is the same reference measure of Figure 1.4b based on the actual distance of the respondent's household income from the average family income. In graph (b), the green dashed line represents the percentile of the respondent's household income (right vertical axis). In graph (c), instead of a measure of income position, the green line represents the income tax burden of the respondent's household (the proportion of income paid in taxes) on the right vertical axis.

in the analysis excludes households above the 90th income percentile, affected by the top coding of the income variable.

$$PerceivedPosition_{it} = \beta_1 \frac{\tau(y_{it})}{r_{K_t}} + \beta_2 \ln(y_{it}) + \beta_3 FinChange_{it} + \Gamma' \text{Controls}_{it} + \alpha_i + \epsilon_{it} \tag{1.10}$$

Results are shown in Table 1.1, columns 1 to 4. The first 3 columns treat the observations as cross-sections, clustering errors by household ID. The fourth column corresponds to the panel regression. The relative tax burden has a highly significant explanatory power for perceived position, and it is robust across specifications. Moreover, its size is not negligible: the coefficient for a percentage point change in relative tax burden is close to the effect of an increase of family income by one per cent. The other four columns (5 to 8) correspond to the same regressions using Q2 as the dependent variable. In this case, the statistical significance of household income practically disappears, while the relative tax burden remains with a very stable coefficient value and very high statistical significance (above 99% confidence level). However, the explanatory power of the regressions for Q2 (measured by the Pseudo R-squared) is significantly lower, reflecting that the ambiguity of the question makes it less related to household income (people may interpret very differently what the social ladder is based on).

Table 1.1: Effect of the relative tax burden on the perceived position

	Q1: Distance from average income				Q2: Position in the social ladder			
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Tax Burden (as % of potential maximum)	0.041***	0.027***	0.037***	0.040***	0.016***	0.013***	0.015***	0.019***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Family Income (log, constant USD 2000)		0.252***	0.050***	0.047***		0.050***	-0.036*	-0.014
		(0.013)	(0.015)	(0.015)		(0.018)	(0.022)	(0.021)
Change in financial situation			0.364***	0.336***			0.132***	0.105***
			(0.008)	(0.007)			(0.014)	(0.013)
Observations	49,379	49,379	49,042	49,217	14,515	14,515	12,333	12,360
Year FE	NO	NO	YES	YES	NO	NO	YES	YES
Demographic Controls	NO	NO	YES	NO	NO	NO	YES	NO
Household FE	NO	NO	NO	YES	NO	NO	NO	YES
Pseudo R-squared	0.0961	0.102	0.138	0.315	0.0122	0.0123	0.0275	0.306
Number of Households				3,904				2,841

Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Demographic controls include age, gender, marital status, and educational level.

1.3.4 Limitations

Although the results of this section are robust to different econometric specifications and measures of the tax burden, they should be taken with caution. This section provides some suggestive evidence of a relation between tax schedules

and perceived distributions, but the data and context used for the analysis suffer severe limitations. First, the income data from the GSS is encoded within income ranges, introducing variation in income that is purely a result of changes in the definition of such ranges with different versions of the survey used in different years. Second, that same limitation of the income variable equally affects the calculation of income taxes, which is also likely to introduce spurious variation in my measure of tax burden. Third, the lack of information on tax deductions reduces the accuracy of the measure of tax burden. Finally, my hypothesis suggests a reciprocal effect of taxpayers electing policies that simultaneously affect their perception and may subsequently change their policy preferences. Since elected politicians introduced the tax changes, those are likely to be endogenous, and it is difficult to identify causality. Moreover, changes in the income tax law were often just a part of broader fiscal policies affecting several taxes, and even within the tax law, they involved changes in the number of bands, thresholds, and tax rates, all at the same time. To accurately understand what elements of the tax schedule are informative to taxpayers, we would need a change that only affected either the tax-free allowance, the other tax band thresholds, or the marginal tax rates, one at a time.

1.4 Experimental approach

To overcome the limitations presented in the previous section, I designed an online experiment to analyse the influence of progressive tax schedules on perceived income distributions. I set up a synthetic environment that tried to mimic a real-life situation: participants had to perform some tasks from which they earned a level of income subject to a tax. After learning their earned income and the tax system in place, participants reported measures of perceived statistics of the income distribution. Participants faced a tax schedule that was randomised across treatment groups. This experiment focused explicitly on the role of tax schedules creating progressivity (differences) in the tax burden, and compared a proportional tax system (same flat tax rate on all incomes) with a very simple progressive tax system with two (increasing) marginal tax rates. Therefore, this experiment is limited to testing the main underlying assumption of the theoretical model: that tax systems directly impact perceived income distributions. Further experiments will be necessary to test the specific theoretical predictions comparing different levels of relative tax burden.

An online experiment provides the opportunity to introduce very specific changes to the tax system applied, one at a time, holding all other variables constant. It also allows controlling the set of information made available to respondents

at every moment, thus ensuring a situation as close as possible to the one assumed in the theoretical model. I programmed the experiment using oTree, an open-source software for experiments and surveys created by Chen et al. (2016) based on Python, and implemented it on Amazon Mechanical Turk with location restricted to USA workers only. Respondents were matched to records of other 99 previous participants, creating an overall group of 100 individuals on which they had to infer different income distribution statistics. Monetary amounts in the experiment are expressed in Experimental Currency Units (ECU) to represent realistic yearly income values closely. The equivalence with real-world currency is 50,000 ECU to 1 USD. All tasks and questions are incentivised, with participants' final payment based on their performance on the tasks and the accuracy of their answers. A random lottery selects which questions to pay, eliciting truthful beliefs by eliminating hedging opportunities. In addition to the amount earned through the experiment, all participants were paid a participation fee of 0.40 USD upon completion.

All components and concepts in the experiment were tested with volunteers of different ages and backgrounds to ensure a correct level of comprehensibility. The experiment received internal ethical approval from the Department of Economics at the University of Warwick and was registered at the American Economic Association's registry for randomized controlled trials.[11] Further technical details, including the set of instructions, explanations of the reward structure chosen to ensure the right incentives, screenshots of different sections and more, can be found in Appendix A.1.

1.4.1 Structure of the experiment

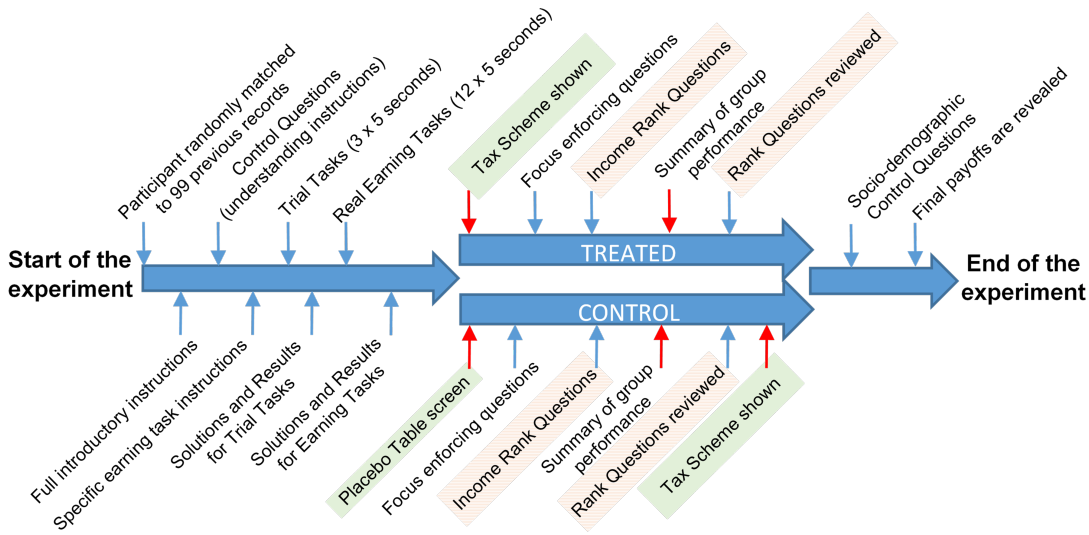
The experiment was implemented in May 2020 and had an average duration of approximately 12 minutes. As summarised by Figure 1.8, the experiment has a first part where participants had to solve twelve tasks to earn their gross income, and they were later given different pieces of information before being asked the target questions on statistics of the perceived distribution. There were two treated groups: one facing a proportional flat tax system and another facing a progressive tax system with two bands. The control group saw a placebo screen displaying a table with similar numbers but framed as a *captcha*[12] to prove they were not a bot and were only revealed the tax system at the end of

[11]The ethical approval reference is *ECONPGR 05/19*, and the AEA RCT ID code is *AEARCTR-0003365*.

[12]Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA).

the experiment. After the first measures of perceived statistics were collected, all participants saw the results of a subset of other players in their group (reference group), and they were given the chance to modify their answers. This provides a measure of elasticity of respondents' beliefs to a set of information that is objectively informative (even if partial), adding validity to the findings of this study. This information was disclosed after treated groups had already learnt the tax system in place and had provided an initial measure of their perceived statistics of the income distribution.

Figure 1.8: Timeline of the experiment



Notes: Randomization is highlighted by the red arrows. Allocation to groups was also random, and the green background boxes highlight the screens that were different for the treatment and control groups. Orange background boxes highlight the main target questions of the experiment.

The questions asked in this experiment were:

1. **Self-perceived rank (Q1):** *How do you think you rank among the 100 players in your group?*

2. **Perceived average income (Q2):** *What do you think was the average income in your group (you included)?*

3. **Perceived position relative to the mean (Q3):** *With what probability do you think your income may be above the group average?*

1.4.2 Identification strategy and randomization

As Figure 1.8 shows, the experiment was designed to compare between individuals in the control group and each of the two treatment groups. Identification was

achieved through random allocation of participants to these three groups:

- **Treatment 1 (T_1):** participants saw the proportional tax schedule before being asked the target questions. The proportional tax system consisted of a unique average tax rate for all levels of income.
- **Treatment 2 (T_2):** participants saw the progressive tax schedule before being asked the target questions. The progressive tax system consisted of a tax-free allowance and a marginal tax rate above that amount.
- **Control:** participants saw a placebo screen instead of the tax schedule, which was presented as a *captcha* to verify they were not bots. They were aware their income would be subject to taxes at the end of the experiment, but they were not told the exact tax system until the end.

To ensure that treated participants paid attention to the tax schedule and control individuals exerted similar effort on the placebo screen, all participants were required to input a few pieces of information. They had to type their gross income earned from the 12 tasks, their proportion of income paid in taxes (control individuals were asked to calculate a ratio of the two values shown in the placebo screen), and the maximum level of income that paid zero taxes (control individuals were asked to input the highest of the values displayed in a table).

Participants were asked to report their beliefs on the statistics Q1, Q2, and Q3 after seeing those tax schedules (or the placebo screen if in the control group). Right after introducing these measures of beliefs, participants saw a little graph showing the income of other nine players in their reference group and were given the opportunity to amend their answers. In this experiment, participants were randomly matched with one of two groups of previous records: high performing or low performing, the latter displaying lower median and average incomes (see Figure 1.11).

Therefore, the measures of perceived statistics on the distribution taken at first (X_{i1}) were based on the participant's own income (y_i, measured in thousand ECU), the range of possible incomes ($M \in [0, 120]$), prior personal knowledge and biases (ϵ_i, unobservable) and, in the case of treated individuals, the information from the tax schedule (T_1 or T_2). As mentioned before, after collecting these measures, participants were shown partial information on their reference group (g_H or g_L) and were allowed to modify their answers, hence reporting a final measure $X_{i2}(y_i, M, T_i, g_i, \epsilon_i)$. Variation between individuals occurs in income (endogenous), tax system (exogenous), and reference group (exogenous).

Figure 1.10 depicts the distribution of tax burdens across income levels (left graph) and the difference between gross and net incomes under each tax system (right graph). Those earning 60,000 ECU were taxed the same under both systems. Remember, participants in the control group did not count with any of this information before reporting their perceived statistics and were only shown the tax system in place and their tax due at the end of the experiment, after all target questions had been answered.

Based on the theoretical framework proposed in Section II of this chapter, taxpayers may derive information about their position from progressive tax schedules, but they cannot do so from proportional tax systems. This is because a proportional flat tax rate implies that everybody pays the same share of income in taxes (and hence there is no difference in tax burden across income levels). Therefore, we would expect to find significant differences between Control and Treatment 2 (progressive tax schedule) but not between Control and Treatment

Figure 1.9: Randomized tax schedules shown to participants

Income (ECU)	ATR*
From 0$	15%

(a) Flat/Proportional Tax (Treatment 1)

Income (ECU)	MTR*
0$ to 40,000$	0%
From 40,001$	45%

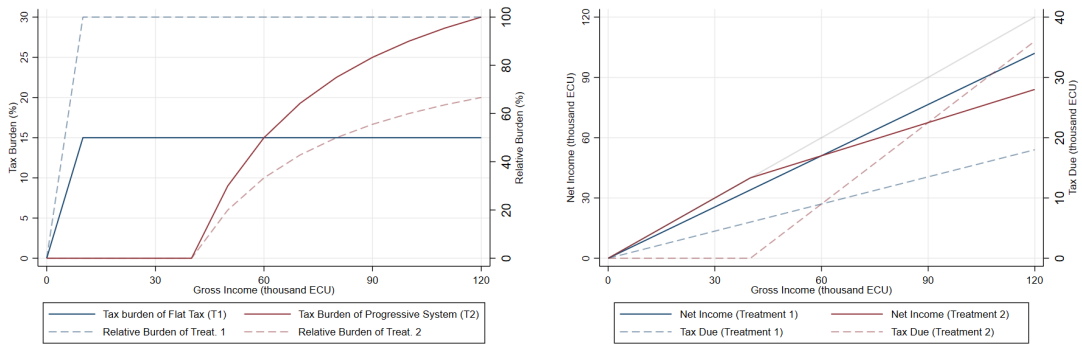
(b) Progressive Tax (Treatment 2)

40,000
10,000

(c) Placebo Captcha (Control)

Notes: ATR stands for Average Tax Rate, applied on all gross income. MTR stands for Marginal Tax Rate, applied only on the fraction of income that falls within each band.

Figure 1.10: Income taxes, by tax schedule



(a) Distribution of tax burdens

(b) Gross vs Net incomes

Notes: Graph (a) shows the tax burden that individuals earning different levels of income faced depending on their tax schedule, as their average tax rate (solid lines, on the left vertical axis), or relative to the top marginal rate (dashed lines, on the right vertical axis). The solid lines in graph (b) illustrate the differences in tax burden by plotting the net income corresponding to each level of gross income under the two possible tax schedules. The dashed lines show the amount of taxes paid by each level of gross income.

1 (proportional tax system). However, if significant differences were identified between Treatment 1 and Control, without significant differences between both treatments, my hypothesis about the *tax burden heuristic* would be rejected while confirming a role of taxes on perceived income distributions through a mechanism to be further explored. On the other hand, if no statistical differences were identified between the Control group and either of the Treatment groups, that would indicate that income taxes are unlikely to be the main driver of the large bias observed at the bottom of the income distribution in the GSS data.

It is important to notice that if participants suffered from representativeness bias[13], the starting bias of those above and below the mean would have the opposite sign, potentially affecting any posterior updating of beliefs. Therefore, I created a dummy (H_i) to differentiate effects on high-income levels with respect to low-income levels. I defined as *top half* incomes from $60,000$ ECU ($H = 1 \text{ if } y_i \geq 60$).

1.4.3 Empirical analysis

The sample of participants in the experiment consisted of $1,535$ people, distributed evenly between the three groups, as seen in Table 1.2. The percentage in parenthesis is the proportion of participants in each cell being matched to the high-performance group. To reduce noise, I exclude the tails of the distribution, thus focusing on the income range $[10, 100]$. This reduces the sample to $1,368$ valid observations.

Table 1.2: Distribution of participants across groups, by income level

	Income (thousand ECU)													
	0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**Total**
Treatment 1	33	40	63	62	71	63	53	52	33	17	14	4	3	508
	(51.5%)	(47.5%)	(57.1%)	(51.6%)	(50.7%)	(46%)	(60.4%)	(51.9%)	(51.5%)	(35.3%)	(57.1%)	(50%)	(66.7%)	(51.8%)
Treatment 2	59	58	36	61	66	50	58	45	55	22	12	5	1	528
	(42.4%)	(50%)	(44.4%)	(54.1%)	(47%)	(46%)	(50%)	(55.6%)	(58.2%)	(63.6%)	(50%)	(0%)	(100%)	(50%)
Control	51	49	39	47	66	60	72	46	31	16	11	10	1	499
	(56.9%)	(51%)	(48.7%)	(46.8%)	(54.5%)	(53.3%)	(52.8%)	(39.1%)	(41.9%)	(56.3%)	(36.4%)	(70%)	(0%)	(50.5%)
Total	143	147	138	170	203	173	183	143	119	55	37	19	5	1535
	(49.7%)	(49.7%)	(51.4%)	(51.2%)	(50.7%)	(48.6%)	(54.1%)	(49%)	(52.1%)	(52.7%)	(48.6%)	(47.4%)	(60%)	(50.7%)

Notes: In each cell, the number on the top is the number of individuals with that income level in a given treatment group. The percentage in parenthesis is the proportion of those who were matched to the high-performing reference group.

For the first measures of perceived statistics of the distribution (X_{i1}), participants only knew the set of tasks, their gross earned income (y_i) and, treated individuals, the tax system (T_1 or T_2). In the regression analysis, income effects are allowed to

[13]The consequence of the representativeness bias in this context would be that agents believe they are more representative of the population than they are, hence underestimating their distance from the average income.

be non-linear by using a set of dummies (C), and the control group is used as the base category, adding dummies for each treatment group. I also add the dummy (H_i) for the top half of the income range included in the analysis ($y_i \geq 60$) interacted with the treatment dummies. For individuals in Treatment 2 (progressive tax schedule), the dummy for the top half of the income distribution also captures the effect of the relative tax burden (higher on the top half). Notice that the income dummies already absorb any differences between the lower and top half of the income range in the control group. The set of socio-demographic controls (W_i) added to the regression are age, education level, previous experience filing taxes, and the quintile of the USA income distribution they believe to belong to in real life. Moreover, the measures of relative position (self-perceived percentile and perceived probability of being above the mean) are likely to be affected by the perceived average income. Therefore, I run an additional specification for those two regressions, controlling for perceived average income ($\tilde{y}_i$). This specification allows testing whether the tax has a direct impact on the perceived position, as suggested by the *tax burden heuristic*, or all the effect goes through an update on the whole perceived distribution (reflected by the change in perceived mean). Continuous variables are used in natural logarithms so that the coefficients can be interpreted as percentage changes. The estimation equation is:

$$\ln X_{i1} = \theta_0 + \Theta_1' \mathbb{1}\{T_i\} + \Theta_2' \mathbb{1}\{T_i \times H_i\} + [\theta_3 \ln \tilde{y}_i] + \Theta_4' \mathrm{C}_i + \Theta_5' \mathrm{W}_i + \varepsilon_i \quad (1.11)$$

Table 1.3 shows the results of the ordinary least squares (OLS) regressions for the first measures of perceived statistics of the income distribution. None of the three measures (Q1, Q2, Q3) shows any statistically significant difference between either of the treated groups and the control group. Nevertheless, the sign of the coefficients is consistent across specifications, and it is the opposite for each treatment (T_1, T_2). Individuals facing the flat tax system seem to perceive themselves at a slightly higher percentile than those in the control group if their income is in the bottom half ($y_i < 60$), and slightly lower if their income is in the top half ($y_i \geq 60$). On the other hand, individuals facing the progressive tax system seem to believe to be at a slightly lower percentile than those in the control group only if their income is below $60,000$ ECU.

Since participants were allowed to modify their answers after seeing the summary of incomes of other members in their group, I ran the same regression with those final *amended* measures, incorporating the new set of information. Given that all three groups (T_1, T_2 and *Control*) saw the partial summary statistics of the income distribution of their reference group at the same point in the experiment,

Table 1.3: Impact of tax schedules on perceived rank and average income (first measure)

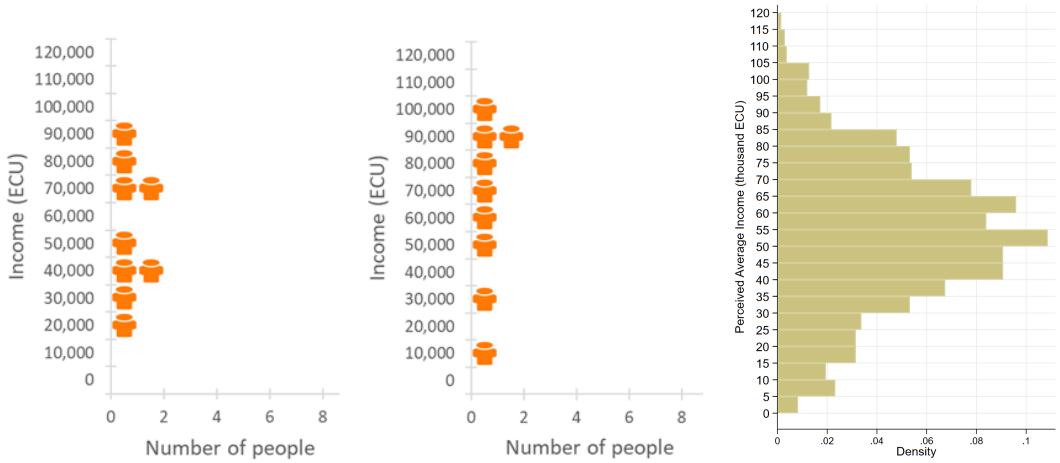
	Q2: Mean Income		Q1: Percentile			Q3: $Pr(y_i > \tilde{y})$		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flat Tax System (T_1)	-0.070	-0.078	0.014	0.024	0.004	0.062	0.078	0.044
	(0.051)	(0.050)	(0.068)	(0.065)	(0.064)	(0.117)	(0.111)	(0.108)
Progressive Tax System (T_2)	0.026	0.021	-0.081	-0.103	-0.098	-0.020	-0.053	-0.044
	(0.049)	(0.048)	(0.073)	(0.070)	(0.069)	(0.124)	(0.116)	(0.114)
Flat Tax × Top Half ($T_1 \times H_i$)	0.064	0.080	-0.080	-0.097	-0.076	-0.100	-0.136	-0.101
	(0.061)	(0.061)	(0.086)	(0.086)	(0.084)	(0.137)	(0.135)	(0.131)
Prog. Tax × Top Half ($T_2 \times H_i$)	-0.007	-0.005	0.052	0.094	0.093	-0.075	0.011	0.009
	(0.058)	(0.059)	(0.089)	(0.088)	(0.086)	(0.147)	(0.142)	(0.140)
Perceived mean income ($\ln \tilde{y}_i$)					-0.259***			-0.437***
					(0.051)			(0.093)
Observations	1,368	1,363	1,368	1,363	1,363	1,368	1,363	1,363
R-squared	0.297	0.306	0.122	0.171	0.198	0.206	0.275	0.301
Income Dummies	YES	YES	YES	YES	YES	YES	YES	YES
Socio-Dem. Controls	NO	YES	NO	YES	YES	NO	YES	YES

Robust standard errors in parenthesis. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. *Notes:* Socio-demographic controls include age, gender, employment status, and educational level. The first two columns report coefficients for the regressions of the perceived average income (Q2). The dependent variable in columns 3 to 5 is the perceived income percentile (Q1). The final three columns correspond to the regressions of the perceived probability of being above the average income (Q3). The coefficients of the income dummies and control variables have been removed from this output table (vectors Θ_4' and Θ_5' in Equation 1.11).

it is impossible to differentiate how much of the change between both measures results merely from realising they had introduced a wrong amount at first, and how much was a result of the disclosed results of other participants. For that reason, the most relevant regression will be one using the final measure only, after participants were allowed to correct their initial answers. Nevertheless, I include the regression on the *correction* (updating) process for robustness.

Participants were shown a summary of the performance of other nine players in their reference group, which could be a low-performing or a high-performing group. The two first graphs from the left in Figure 1.11 correspond to those pieces of information participants received. The low-performing reference (left) has the income level of nine players, with a median value $50,000$ and mean $54,444$ ECU. The summary of the high-performing group (centre), on the other hand, has a median value $70,000$ and mean $64,444$ ECU. The last graph in the same figure shows the distribution of the reported perceived average income in the first set of questions (before seeing the group information). Answers follow a normal distribution centred around $52,500$ ECU. Fifty per cent of respondents reported a prior belief for the average income below the mean of the lower reference group, while the number of participants with a prior belief below the mean income of the higher reference group was substantially larger, up to seventy per cent.

Figure 1.11: Reference groups and prior perceived average income



(a) Lower Group ($R_i = 0$) (b) Higher Group ($R_i = 1$) (c) Distribution of perceptions

Notes: The partial distributions in (a) and (b) are the pieces of information about other participants' performance that players were shown after submitting their initial set of answers. Chart (c), on the other hand, summarises the distribution of initial answers about the perceived average income, with mean value at 37.6 thousand ECU.

Participants were given a chance to modify their initial answers after seeing this new piece of information. If participants were to update their beliefs merely based on the new piece of information, and given the distribution of initial answers (with an average estimated mean income below 40,000 ECU), we would expect a statistically significant positive change in perceived mean (Q2) for respondents matched to either group, though larger for those with the higher reference group ($R_i = 1$), with no differences between control and treated groups. The sign of the coefficients should be the exact opposite for the measures of perceived position (Q1 and Q3).

Nevertheless, I already mentioned that giving participants the opportunity to correct their answers may cause changes that respond to other factors than the new piece of information. Most of them are unobservable, but I can control for the information they saw previously (treatment group). Therefore, I regress the changes to the reported measures, denoted by $\triangle X_i = \ln(\frac{X_{i2}}{X_{i1}})$. As defined in the estimation equation 1.12, I control for mean reversion by adding the initial measure X_{i1} in natural logarithms, add treatment dummies for changes uncorrelated with the reference group, and interact the treatment and reference group dummies with the dummy for the top half values of the income range (H_i) to allow for differential effects on high and low incomes. To identify the direct impact of tax systems on perceived rank, I control for changes in perceived average income too, $\triangle \tilde{y}_i = \ln(\frac{\tilde{y}_{i2}}{\tilde{y}_{i1}})$. As in Equation 1.3, C_i is the set of income dummies, and W_i

the set of controls. Results of this set of OLS regressions are shown in Table 1.4.

$$\triangle X_i = \gamma_0 + \gamma_1 \ln X_{i1} + \Gamma_2' \mathbb{1}\{T_i\} + \Gamma_3' \mathbb{1}\{T_i \times H_i\} + \gamma_4 R_i + \gamma_5 \{R_i \times H_i\} + [\gamma_6 \triangle \tilde{y}_i] + \Gamma_7' \mathrm{C}_i + \Gamma_8' \mathrm{W}_i + \varepsilon_i \quad (1.12)$$

Table 1.4: Impact of reference groups on perceived rank and average income (change)

	Q2: $\triangle$ Mean Income		Q1: $\triangle$Percentile			Q3: $\triangle Pr(y_i > \tilde{y})$		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Prior Belief	-0.258***	-0.266***	-0.207***	-0.217***	-0.202***	-0.197***	-0.219***	-0.204***
	(0.045)	(0.045)	(0.024)	(0.024)	(0.024)	(0.026)	(0.027)	(0.027)
Flat Tax System (T_1)	0.096***	0.093***	-0.052	-0.052	-0.016	-0.169**	-0.164**	-0.100
	(0.026)	(0.026)	(0.042)	(0.042)	(0.040)	(0.069)	(0.068)	(0.071)
Progressive Tax System (T_2)	0.103***	0.102***	-0.012	-0.010	0.022	-0.243***	-0.246***	-0.190**
	(0.027)	(0.026)	(0.044)	(0.043)	(0.042)	(0.078)	(0.078)	(0.079)
Flat Tax $\times$ Top Half ($T_1 \times H_i$)	-0.107***	-0.108***	0.070	0.072	0.031	0.142	0.139	0.066
	(0.031)	(0.032)	(0.053)	(0.053)	(0.051)	(0.087)	(0.087)	(0.089)
Prog. Tax $\times$ Top Half ($T_2 \times H_i$)	-0.136***	-0.141***	0.045	0.048	0.001	0.157	0.171*	0.091
	(0.031)	(0.032)	(0.054)	(0.053)	(0.052)	(0.101)	(0.102)	(0.103)
Higher Reference Group (R_i)	0.055**	0.054**	-0.111***	-0.117***	-0.100***	-0.096	-0.104	-0.074
	(0.022)	(0.022)	(0.037)	(0.037)	(0.036)	(0.064)	(0.063)	(0.061)
Higher Ref. Group $\times$ Top Half ($R_i \times H_i$)	0.103***	0.098***	-0.049*	-0.040	-0.007	-0.185***	-0.179***	-0.118**
	(0.014)	(0.014)	(0.028)	(0.028)	(0.029)	(0.049)	(0.050)	(0.050)
Change of perceived mean income ($\triangle\tilde{y}_i$)					-0.322***			-0.574***
					(0.068)			(0.150)
Observations	1,368	1,363	1,368	1,363	1,363	1,368	1,363	1,363
R-squared	0.259	0.274	0.156	0.164	0.201	0.126	0.142	0.181
Income Dummies	YES	YES	YES	YES	YES	YES	YES	YES
Socio-Dem. Controls	NO	YES	NO	YES	YES	NO	YES	YES

Robust standard errors in parenthesis. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.
Notes: The regressions in this table have the change in the reported statistics as dependent variables. That is, the log-difference between the initial value reported and the final value submitted when individuals were offered the opportunity to amend their answers after seeing the scores of a subset of other 9 participants in their group. The first two columns report coefficients for the regressions on the change in perceived average income (Q2). The dependent variable in columns 3 to 5 is the change in perceived income percentile (Q1). The final three columns correspond to the regressions of the change in perceived probability of being above the average income (Q3). Socio-demographic controls include age, gender, employment status, and educational level.

Individuals in the lower half of incomes of both treated groups increased their perceived average income (columns 1 and 2) by around 10% more than control individuals. Among individuals in the higher half of incomes, those facing the progressive system (T_2) reduced their perceived mean by a statistically significant 3% compared to the control group, while there weren't significant differences between those facing the proportional system (T_1) and the control group. Moreover, individuals seeing the summary for the higher-performing reference group significantly increased their perceived average income (Q2) by another five per cent, and the magnitude of such increase was 10 percentage points larger for those in the top half of the income range (see interaction term $R_i \times H_i$). Seeing summary information from the higher-performing reference group also had a negative and

statistically significant impact on the perceived percentile (Q1) and the probability of being above the mean (Q3). Changes in perceived percentile, however, do not significantly differ between control and treated individuals. Significant differences between Control and Treated groups arise in the other measure of position (Q3). Individuals facing the flat tax system (T_1) reduced their perceived probability of being above the mean, although its significance level drops after controlling for the actual update on the perceived average income level. Individuals facing the progressive tax system (T_2), on the other hand, also significantly reduced their perceived probability to be above the mean, and such change remains statistically significant after controlling for the actual change in perceived average income level (column 8). This means that the change in perceived position of those in Treatment 2 went beyond what can be explained through the change in perceived average income.

The last set of answers is likely to report the respondent's beliefs more accurately, and having a richer set of information is likely to reduce variance. Therefore, the last regression is the most relevant to confirming tax schedules' influence on perceived income distributions and perceived position. Equation 1.13 aims to identify the causal impact of tax systems and reference groups on perceived average income, perceived percentile, and perceived position relative to the mean:

$$\begin{aligned} X_{i2} = \lambda_0 + \Lambda_1' \mathbb{1}\{T_i\} + \Lambda_2' \mathbb{1}\{T_i \times H_i\} + \lambda_3 \mathbb{1}\{R_i\} \\ + \lambda_4 \mathbb{1}\{R_i \times H_i\} + [\lambda_5 \ln \tilde{y}_i] + \Lambda_6' \mathrm{C}_i + \Lambda_7' \mathrm{W}_i + \epsilon_i \end{aligned} \quad (1.13)$$

Effectively, Table 1.5 confirms a significant impact of the progressive tax system. First, columns 1 and 2 (Q2) confirm that the progressive tax system had a significant impact on the average income perceived by lower-income individuals (most of which are within the tax-free allowance), while the flat tax system did not generate significant differences in Q2 compared to the control group. Individuals with incomes below 60,000 ECU in Treatment 2 (progressive tax) perceived an average income level in their reference group 12% larger than those in the Control group and 8% larger than those in the Treatment 1 group (flat tax). Columns 6 to 8 also confirm that the progressive system had a statistically significant impact on the perceived position with respect to the average income, which is the most relevant information according to classical models of preferences for redistribution, where individuals below the average income have incentives to support redistributive policies since they would benefit from them. Compared to

control individuals, those facing the progressive tax system believed they were nearly 25% less likely to be above the mean if their income was in the bottom half, although the impact is only half that magnitude on individuals with higher incomes ($y_i \geq 60,000$ ECU), controlling for socio-demographic characteristics. The impact of the progressive tax system on the perceived probability of being above the mean remains above 90% confidence level after controlling for the perceived average income level, which could indicate taxpayers assessed their position relative to the average income by using the *tax burden heuristic*. Columns 3 to 5, however, do not show statistical significance for either treatment effect, and therefore the regression fails to reject that the tax system had no impact in shaping the perceived income rank (percentile) of taxpayers.

Table 1.5: Impact of tax schedules on perceived rank and average income (final measure)

	Q2: Mean		Q1: Percentile			Q3: $Pr(y_i > \tilde{y})$		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Flat Tax System (T_1)	0.045	0.036	-0.041	-0.033	-0.023	-0.119	-0.103	-0.083
	(0.048)	(0.046)	(0.069)	(0.067)	(0.065)	(0.121)	(0.114)	(0.111)
Progressive Tax System (T_2)	0.123***	0.117***	-0.077	-0.091	-0.058	-0.259**	-0.287**	-0.223*
	(0.046)	(0.044)	(0.075)	(0.071)	(0.071)	(0.128)	(0.119)	(0.117)
Flat Tax × Top Half ($T_1 \times H_i$)	-0.058	-0.048	0.007	-0.005	-0.018	0.064	0.034	0.008
	(0.057)	(0.057)	(0.086)	(0.086)	(0.084)	(0.144)	(0.140)	(0.136)
Prog. Tax × Top Half ($T_2 \times H_i$)	-0.139**	-0.143***	0.087	0.121	0.081	0.100	0.181	0.104
	(0.055)	(0.054)	(0.090)	(0.088)	(0.088)	(0.153)	(0.149)	(0.146)
Higher Reference Group (R_i)	0.058	0.058	-0.122**	-0.123**	-0.107*	-0.080	-0.088	-0.057
	(0.037)	(0.036)	(0.059)	(0.057)	(0.056)	(0.101)	(0.095)	(0.094)
Top Half × Higher Ref. Group ($H_i \times R_i$)	0.085***	0.080***	-0.058	-0.035	-0.012	-0.223***	-0.195***	-0.152**
	(0.024)	(0.025)	(0.042)	(0.044)	(0.044)	(0.066)	(0.071)	(0.070)
Perceived mean income ($\ln \tilde{y}_i$)					-0.282***			-0.544***
					(0.060)			(0.096)
Observations	1,368	1,363	1,368	1,363	1,363	1,368	1,363	1,363
R-squared	0.239	0.251	0.208	0.250	0.274	0.271	0.335	0.362
Income Dummies	YES	YES	YES	YES	YES	YES	YES	YES
Socio-Dem. Controls	NO	YES	NO	YES	YES	NO	YES	YES

Robust standard errors in parenthesis. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.
Notes: The first two columns report coefficients for the regressions on the perceived average income (Q2). The dependent variable in columns 3 to 5 is the respondents' perceived income percentile (Q1). The final three columns correspond to the regressions of the perceived probability of being above the average income (Q3). Socio-demographic controls include age, gender, employment status, and educational level. The coefficients of the control variables and the income dummies are not included in this table.

1.5 Conclusion

This research project aims to fill a relevant existing gap in the literature of redistributive preferences. Its initial results, despite its limitations, seem to demonstrate that tax systems play a role in shaping perceived income distributions and, as a result, may influence support for redistributive policies by biasing taxpayer perceptions.

The initial exploratory analysis using the longitudinal survey GSS in the United States and exploiting quasi-experimental variation from tax law changes shows a systematic inflexion in perceived position for households with incomes below the tax-free allowance. This paper presents a rational framework to explain why taxpayers may infer signals from tax systems using the most salient information: their own tax burden and the maximum possible tax rate in the system. A similar reasoning could be used to develop a model with different reference points. For example, one where taxpayers at the bottom of the distribution used the tax-free allowance as a reference, while taxpayers at the top of the distribution used the top threshold as a reference.

Despite requiring large sample sizes to ensure enough statistical power to identify small effects, online experiments provide a synthetic environment where the different key elements of tax schedules can be tested one at a time. My experiment, using a large number of participants to compare two very different tax systems, revealed statistically significant differences in perceived income distributions of individuals facing a proportional flat tax system compared to those facing a progressive tax system. On average, individuals facing the progressive tax system believe they are between 10% and 25% less likely to be above the average income than those in the control and flat tax system groups. Moreover, the effect remains significant at the 10% confidence level and with a fairly stable value even after controlling for differences in the perceived average income value. At the same time, the perceived average income is significantly higher (+12%) for low-income individuals facing the progressive tax system than for those in the control and flat tax groups. These findings seem to indicate that tax systems influence perceptions of relative position as well as overall distributions of income. On the other hand, however, differences in perceived percentile were not significant, hence challenging the robustness of the findings.

Despite its limitations, the conclusions of this paper are of relevance to the literature in redistributive preferences, contributing to solving the *inequality-redistribution puzzle*. My results suggest that taxpayers use tax information to

create their beliefs on income distributions and their position within those. Further research is needed to gain a better understanding of the mechanism behind the impact of progressive tax systems on perceived income distributions, as well as the role of income tax credits and different tax rebates, which were not considered in this study. If posterior studies confirmed the existence of the proposed tax burden heuristic, it would highlight the risks of misinformed voters. If tax systems themselves contribute to biasing perceptions of citizens, it is of utmost importance to provide accurate information to taxpayers on their true position in the distribution so they can correctly assess what policies would be of most benefit to them. Otherwise, by reducing the highest marginal tax rates as it happened during the second half of the 20th century, taxpayers at the lower end of the income distribution saw their tax burden get closer to the maximum tax rate, potentially inducing a biased overestimation of their position relative to the average family income. This, at the same time, could translate into lower support for redistributive policies, hence increasing support for the tax reduction on high incomes that caused the bias in the first place, and leading to a continued reduction of progressivity in income tax schedules.

2 Tax incidence in online markets: Pass-through of the sugar tax on Amazon UK

2.1 Introduction

According to NHS data, nearly two-thirds of adults in England were overweight or obese in 2018, a condition directly accountable for 11,117 hospital admissions in that year (Health and Social Care Information Centre, 2020). It is estimated that overweight and obesity-related health conditions cost the NHS over six billion pounds per year (Public Health England, 2017). Obesity is not only responsible for thousands of deaths each year but also increases the risk of developing certain types of cancer, heart diseases, and type 2 diabetes. Consumption of sugary drinks has been found to be highly correlated with the increase in obesity experienced by western economies in the last decades and is directly associated with several chronic diseases (WHO, 2015b). To tackle this, the World Health Organisation issued recommendations on the use of fiscal policies to increase the prices of sugary drinks and reduce their current consumption levels (WHO, 2015a, 2016). In addition to collecting revenues that can be spent on public health policies, an argument in favour of taxing beverages with high sugar content is that people do not internalise the health costs (to themselves and to the health system) of consuming unhealthy drinks. By introducing a tax that accounts for such externalities, consumers' final purchase decisions would be closer to the social optimum (Allcott et al., 2019b).

According to basic microeconomic theory, prices should increase in response to the introduction of a tax and, consequently, consumption would decrease. The level to which prices and consumption respond to the new tax depends on the elasticities of demand and supply and the degree of market concentration. It is important to notice that those variables differ across purchasing channels, types of drinks, and package sizes. In a perfectly competitive setting, prices should change by an amount smaller than the tax or, in the presence of constant marginal costs, equal to it (full pass-through). In imperfectly competitive markets, however, taxes could be over-shifted as well as partially passed through to prices, depending on the curvature of the demand function and the level of market concentration (Weyl and Fabinger, 2013). Overall, higher elasticity of demand and higher competition reduce pass-through, while higher elasticity of supply increases it. Finally, the impact on prices and consumption will also depend on the type of tax:

an ad-valorem tax is less likely to be over-shifted than an excise tax (Delipalla and Keen, 1992; Bonnet and Réquillart, 2013).

In the last years, a growing number of countries introduced taxes targeting unhealthy drinks.[1] In Europe, most countries opted for a volumetric excise tax charged per litre, with the exception of Spain, where the government increased the value-added tax (VAT) on all sweetened drinks in January 2021. Belgium, Estonia, Finland[2], Latvia, Norway, and Portugal implemented excise taxes on both sugar and artificially-sweetened beverages, while Hungary, Ireland, the Spanish region of Catalonia, and the United Kingdom (UK) adopted excise taxes only affecting sugar-sweetened beverages (SSB).

Under the name Soft Drinks Industry Levy (SDIL), the UK introduced a tax on soft drinks containing more sugar than 5 grams (g) per 100 millilitres (ml) on the 6th of April 2018, with a higher tax on beverages with sugar content above 8g/100ml, similar to the tiered systems implemented in Catalonia, Estonia, and Ireland. The policy was conceived as an incentive to encourage producers to reduce the sugar contents of their drinks. The new law was announced with two years notice to allow manufacturers enough time to reformulate their beverages and avoid being subject to the tax. This was in contrast with the approach of many other countries that relied mainly on the expected effect of price changes on consumption. Although many producers did indeed reformulate their drinks to reduce sugar contents, the two market leaders (Coke and Pepsi) did not, which emphasises the importance of pass-through as a price mechanism to disincentivise consumption (Bandy et al., 2020).

Given that the theoretical literature on tax pass-through describes scenarios under which under-shifting and over-shifting may occur, the actual proportion of taxes on sugary drinks that is passed on to consumer prices remains an empirical question. Moreover, the level of pass-through is a key statistic necessary to make projections on expected changes in consumption and tax revenues prior to implementation, and one that should be incorporated in optimal tax calculations that so far assume full pass-through (Allcott et al., 2019b). Similarly, simulation studies using structural demand models to predict the impact of the British SDIL in the UK have often assumed that taxes would be fully passed on to prices (Briggs et al., 2013; Tiffin et al., 2015; Dubois et al., 2020). However, the only two existing studies empirically evaluating the effect of the SDIL on prices ex-post have very diverging results. Scarborough et al. (2020) used data from on-the-go pur-

[1] A summary table containing a full list of countries that introduced taxes on soft drinks before February 2021 can be found in Appendix B.1, TableB.1.

[2] The tax on soft drinks in Finland was removed in 2017.

chases and found that prices of drinks in the higher tier only increased by around half the amount of the tax, while those in the lower tier did not change prices at all. O'Connell and Smith (2020), on the other hand, used prices listed by the biggest supermarket chains and found a full pass-through on sugary drinks in both tiers. Evaluations of the soda tax in France also yielded estimates ranging from a very low pass-through of around 39% (Etilé et al., 2018) to a full pass-through (Berardi et al., 2016; Capacci et al., 2019). Studies on Mexico's excise tax on sugary drinks introduced in 2014 seemed to more consistently find over-shifting of the tax on prices of carbonated drinks with pass-through estimates between 112% and 140%, and under-shifting on non-carbonated drinks (Colchero et al., 2015; Grogger, 2017; Campos-Vázquez and Medina-Cortina, 2019). Such heterogeneity across product categories, with carbonated SSB experiencing a larger pass-through than non-carbonated SSB, was also found in France and Berkeley, USA (Falbe et al., 2015; Berardi et al., 2016; Silver et al., 2017; Capacci et al., 2019). Other studies in the United States of America (USA) also identified differences when the tax only applies to a small jurisdiction like Berkeley, resulting in a lower pass-through than when the affected area is larger like the states of Washington and Philadelphia, highlighting the importance of cross-border competition (Rojas and Wang, 2017; Bollinger and Sexton, 2018; Cawley et al., 2020; Seiler et al., 2021). Finally, pass-through has also been found to vary depending on container size, with prices of single-serving sizes increasing by a larger fraction of the tax than those of family-size containers (Colchero et al., 2015; Cawley and Frisvold, 2017; Powell et al., 2020).

To the best of my knowledge, this is the first paper estimating the pass-through of sugar taxes specifically in the online retail market, which accounts for an increasing share of the groceries market and has specific features that can result in different tax pass-through compared to physical stores. On the one hand, online consumers are more likely to be middle-aged and higher income, groups that tend to exhibit lower price elasticity. On the other hand, comparing prices between stores is less costly from a computer browser than walking from one shop to another. However, entry barriers for new sellers may be higher in digital markets than offline since smaller shops and producers are very unlikely to show in the top results of search engines, resulting in more concentrated markets. Learning about the impact of these features on purchasing behaviour is of increasing relevance as online channels expand their market share.

According to ONS data[3], internet food retail sales in the United Kingdom in-

[3]Retail Sales Index internet sales monthly series published by the Office of National Statistics (ONS).

creased significantly over the last 12 years, and the COVID-19 pandemic substantially boosted those numbers. Already before February 2020, internet sales value accounted for 20% of all retailing (excluding automotive fuel), with a quarter of it belonging to predominantly food stores. The UK is the leading country in online grocery sales in the European Union[4], which makes it the best case to separately analyse the specific responses of online markets to tax changes. A priori, the expected differences in tax pass-through between online and traditional retailers are not obvious. On one side, the profile of people most likely to purchase groceries online (25-44 years old and higher income) usually exhibits lower price elasticity than consumers of other age groups or lower income. On the other hand, online search costs are lower than offline, making it easier for the consumer to compare prices between suppliers and thus increasing demand elasticity, even though the number of suppliers online is limited, with only the biggest grocery chains offering this service (higher market concentration).

Using web-scraped daily prices for beverages listed on Amazon UK during the year prior and one year after the introduction of the Soft Drinks Industry Levy in the UK, I created a panel with 134 items, including water, carbonated SSB, sugar-free alternatives, milk and juices. Using milk and water as control groups, I estimate a full pass-through on prices of beverages affected by the tax and some general equilibrium effects that spill over their direct sugar-free alternatives. Furthermore, I analyse heterogeneity between three categories affected by the tax - cola drinks, energy drinks and other flavoured sodas - which differ in their level of consumer price elasticity and market concentration. The estimated pass-through is higher for the categories of drinks facing less elastic demand and with a higher market concentration, as predicted by the theoretical model. Over-shifting is observed for cola-flavoured drinks, while prices of energy drinks and other flavoured sodas experience a lower pass-through. I also explore heterogeneity across sizes and find that the sugar tax was not as effective in increasing prices of the biggest container sizes (1 litre or more). Medium sizes (between 360ml and 999ml) exhibit the largest pass-through.

This paper contributes to three main strands of research. First, to the growing literature evaluating the use of fiscal policies to reduce non-communicable diseases. This is the first study to explicitly evaluate the pass-through of a sugar tax using a unique dataset of prices from a purely online retailer. My findings corroborate those of O'Connell and Smith (2020) regarding the pass-through of

[4]According to the Kantar Worldpanel, online grocery sales in the UK account for 7.2% of the global online grocery sales value in the European Union for 2017/2018. Data accessible at: *https://www.statista.com/statistics/614717*.

the SDIL in the UK and swell the list of papers identifying full pass-through of sugar taxes to prices of carbonated drinks, with over-shifting on specific categories and container sizes. Second, the link between the results of this paper and the theoretical predictions of a model with differentiated products under oligopoly is of relevance to the literature of industrial organisation analysing pass-through under imperfect competition with untaxed substitute goods. Third, the analysis of a purely online retailer and the comparison of my results to those from similar studies using data from retailers with brick-and-mortar stores contributes to the strand of research documenting differences between online and offline pricing strategies. Such comparison provides relevant insights on how pricing strategies of firms selling through both channels may change as the share of their revenues coming from online sales increases.

The remainder of the paper is structured as follows: Section II describes the introduction of the British Soft Drinks Industry Levy in detail; Section III reviews the theory of tax incidence and discusses the differences between online and offline markets that could affect pass-through; Section IV provides detail on the data used and the empirical analysis; Section V concludes.

2.2 The Soft Drinks Industry Levy

The Soft Drinks Industry Levy (SDIL) is a volumetric excise tax levied on soft drinks containing at least 5 grams (g) of sugar per 100 millilitres (ml). It came into effect on 6th April 2018 and excluded drinks with $\geq$ 75% of milk, milk replacements, alcohol replacements, pure fruit juices, liquid flavouring, powder drinks, drinks prepared in opened containers, baby foods, and total diet replacements. Drinks made by small producers were also exempt.[5] The amount of the tax per litre of drink is 18 pennies on beverages containing between 5 and 8 grams of sugar per 100 millilitres, and 24 pennies on drinks with 8 grams or more of sugar per 100ml, incremented by the proportional 20 per cent of Value Added Tax (VAT). This results in 21.6 pennies per litre on beverages in the lower tier and 28.8 pennies per litre on drinks in the higher tier. It is equivalent to a tax of up to 4.32 sterling pounds per kilogram of sugar.

The then chancellor George Osborne announced this policy in March 2016, giving producers two years to reformulate their drinks in order to avoid the tax. On the news release on 5th April 2018 announcing the levy was coming into effect the day after, the HM Treasury reported that 50 per cent of manufacturers had

[5]More details on the SDIL can be found online at *https://www.gov.uk/topic/business-tax/soft-drinks-industry-levy* [last accessed on 17th July 2021].

Table 2.1: Tiers of the Soft Drinks Industry Levy

	Sugar content (per 100 ml)	SDIL (per litre)	VAT (+20%)	**Total Tax (pennies / litre)**	Total Tax on a can of 330ml
Tax free	$< 5g$	$0\ p$	$0p$	$0p$	$0p$
Lower tier	$5g$ to $8g$	$18p$	$3.6p$	$21.6p$	$7p$
Higher tier	$\geq 8g$	$24p$	$4.8p$	$28.8p$	$9p$

Notes: All monetary amounts are in pennies (cents of one sterling pound).

already reduced the sugar content of their drinks.[6] Scarborough et al. (2020) also reported a relevant reduction in the number of drinks with sugar levels above the tax threshold between the time of announcement and April 2018, and even a larger reduction after the tax came into effect. The proportion of soft drinks over 5g of sugar per 100ml in their dataset decreased from above 50 per cent in September 2015 to less than 20 per cent in February 2019, while the proportion of soft drinks in their control group remained fairly stable. Bandy et al. (2020) also found that by 2018 six of the top ten soft drinks producers had modified the sugar content of 50 per cent of their products initially above the threshold set by the SDIL. However, the two leading brands, Coke and Pepsi, did not reformulate their best selling drinks, and neither did Red Bull, which took the lead in the category of energy drinks after Lucozade experienced a severe drop in sales following reformulation to cut the sugar content of its drinks.[7] Despite the reformulation of many sugar-sweetened beverages (SSB), more than a quarter of all available products in the two most popular carbonated categories (cola drinks and energy drinks) still contain high sugar levels.[8]

Another relevant aspect of the levy that differs from other countries' is that the tax becomes payable at the point after production when it is made available for sale (or free of charge) for the first time, rather than at the retail sale to final consumers. For drinks packaged in the UK, the packager must report and pay the levy, while if the drinks are imported from abroad, the responsibility falls on the importer. Therefore, drinks that were packaged or brought to the UK before the 6th of April 2018 are not liable. This means that sellers holding larger stocks or slower stock rotation of liable products may have been able to sell tax-free for some time until they had to refill their stock. As a consequence, pass-through of the tax may not be fully observed until a few weeks or months later. Moreover, the fact that the tax is levied on manufacturers implies that pass-through occurs

[6]The full news release can be found on *https://www.gov.uk/government/news/soft-drinks-industry-levy-comes-into-effect* [last accessed on 24th August 2020].

[7]Data from Kantar Media Computer-assisted web interviews 2018-2020. Sales of Lucozade recovered in subsequent years.

[8]Information extracted from the annual Soft Drinks Review (2018) published by Britvic.

in two steps, from manufacturer to retailer and from retailer to final consumer, introducing double marginalisation due to double mark-up adjustment, which is predicted to decrease pass-through compared to a market with vertical integration or a tax levied directly on consumers (Hong and Li, 2017).

2.3 Theoretical framework

The literature on sugar tax incidence has most often focused on partial equilibrium analysis, disregarding general equilibrium considerations. While taxes on non-alcoholic beverages only affect a tiny fraction of total household expenditure and are unlikely to have significant income effects, there is some evidence of substitution and complementarity between sugar-sweetened drinks and other beverages and food (Allcott et al., 2019a). Therefore, neglecting the general equilibrium effects of the tax may lead to biased estimations of consumer welfare (Kotlikoff and Summers, 1987; Goulder and Williams III, 2003). In this section, I revisit the most relevant theoretical models of tax pass-through and discuss the specific characteristics of online grocery shopping that can influence the proportion of the tax that is passed through to prices.

2.3.1 Partial equilibrium under perfect competition

In the simplest partial equilibrium setting, following Jenkin (1872) for tax incidence analysis under perfect competition (c), one can reach an expression for tax (τ) pass-through (ρ) to prices (p) as a function of the elasticity of demand (η_D) and elasticity of supply (η_S).

Consider a tax (τ) levied on sellers, as in the case of the SDIL. The new equilibrium condition will be $D(p) = S(p - \tau)$, where D and S are the demand and supply functions, respectively. Differentiating with respect to the tax and rearranging[9] yields:

$$\rho_c = \frac{dp}{d\tau} = \frac{\eta_S}{\eta_S + \eta_D} = \frac{1}{1 + \frac{\eta_D}{\eta_S}} \tag{2.1}$$

Therefore, in perfectly competitive markets, pass-through will depend uniquely on the ratio between the elasticities of demand and supply. A perfectly elastic demand ($\eta_D \to \infty$) will yield zero pass-through (as with a perfectly inelastic supply), while a perfectly elastic supply ($\eta_S \to \infty$) would result in a full (100%)

[9]Using the equivalences $\eta_D \equiv -(D'p/Q)$ and $\eta_S \equiv S'p/Q$. Note that for the entirety of this section, η_D defines the elasticity of demand in positive terms.

pass-through of the tax to prices (as with a perfectly inelastic demand). Any combination of imperfectly elastic supply and demand curves will yield a partial tax pass-through (between zero and one), with a larger burden falling on the side with the lowest elasticity.

2.3.2 Partial equilibrium under imperfect competition

In the presence of imperfect competition (m), convex demand curves can produce a pass-through greater than 100%, which is known as over-shifting ($\rho > 1$).

Following Weyl and Fabinger (2013) and building on the analysis of pass-through under a monopoly by Bulow and Pfleiderer (1983), one can reach the expression for pass-through under monopolist supply as in Equation 2.2, where η_{ms} represents the elasticity of the inverse marginal surplus function (a measure of the curvature of demand):

$$\rho_m = \frac{1}{1 + \frac{\eta_D - 1}{\eta_S} + \frac{1}{\eta_{ms}}} \tag{2.2}$$

The curvature of demand will be log-convex if $\eta_{ms} < 0$ and log-concave if $\eta_{ms} > 0$. For the specific case of constant elasticity of demand, η_{ms} equals $-\eta_D$ and, therefore, one can reach a new expression that has two additional terms compared to Equation 2.1, which reduce the size of the denominator:

$$\rho_m = \frac{1}{1 + \frac{\eta_D - 1}{\eta_S} - \frac{1}{\eta_D}} = \frac{1}{1 + \frac{\eta_D}{\eta_S} - \frac{1}{\eta_S} - \frac{1}{\eta_D}} \tag{2.3}$$

Since η_D is always greater than 1 at the optimum under monopoly, $(\eta_D - 1)/\eta_S$ is positive, and there will be overshifting whenever $\eta_S > \eta_D(\eta_D - 1)$. Furthermore, in the presence of constant marginal costs ($\eta_S \to \infty$), Equation 2.3 would be simplified to $\eta_D/(\eta_D - 1)$, which will always be greater than one (overshifting), given demand is elastic ($\eta_D > 1$).

Weyl and Fabinger (2013) further developed a general model allowing for asymmetric, imperfectly competitive firms. In such a model (Equation 2.4), a *conduct parameter* θ is introduced to account for market concentration, based on the Lerner Index.[10] Under many competition models such as Cournot or Bertrand,

[10]The conduct parameter defined by the authors builds on Lerner (1934) and adjusts for the elasticity of demand: $\theta = [(p - mc - \tau)/p]\eta_D$, where mc is the marginal cost.

market concentration θ is a constant (independent of quantity q), and therefore, the second term in the denominator of Equation 2.4 can be dropped.

$$\rho = \frac{1}{1 + \frac{d\theta}{dq}q + \frac{\eta_D - \theta}{\eta_S} + \frac{\theta}{\eta_{ms}}} \tag{2.4}$$

Focusing on the case of constant elasticity of demand and assuming θ independent of q, the formula for pass-through can therefore simplify to the following equation, which again has two additional terms compared to Equation 2.1, reducing the denominator if $\theta > 0$:

$$\rho = \frac{1}{1 + \frac{\eta_D - \theta}{\eta_S} - \frac{\theta}{\eta_D}} = \frac{1}{1 + \frac{\eta_D}{\eta_S} - \frac{\theta}{\eta_S} - \frac{\theta}{\eta_D}} \tag{2.5}$$

From this general Equation 2.5 under the assumption of constant elasticity of demand, one can easily derive the monopolistic case with total market concentration ($\theta = 1$)[11] as in Equation 2.3 and the perfectly competitive case ($\theta = 0$) of Equation 2.1. Also, notice that if marginal costs are constant ($\eta_S = \infty$), there will be overshifting whenever equilibrium prices are above marginal costs.

Using available estimates from the literature and the Weyl and Fabinger (2013) model presented in this section (Equation 2.5), I calculate the elasticity of supply that would be necessary to explain a full pass-through in the case of the UK. According to the Competition Commission report on the AG Barr / Britvic merger inquiry in 2013, the gross margins in this market fluctuate between 16 and 70 per cent (C. C. Inquiry Group, 2013). O'Connell and Smith (2020) estimated a Lerner index of 0.46 for SSB and 0.40 for substitute products using data from the major supermarket chains in the UK. Regarding demand elasticity (η_D) of sugar-sweetened drinks in the UK, Briggs et al. (2013) estimated $\eta_D = 0.81$, while Dubois et al. (2020) estimated elasticities as high as $\eta_D = 2.8$ for some brands included in my analysis, similar to the elasticity estimates of O'Connell and Smith (2020). Therefore, I calculate the elasticities of supply that could explain full pass-through according to the general model of Weyl and Fabinger (2013), with price mark-ups at the lower end of the existing estimates (16% and 46%), and levels of elasticity of demand in line with the findings of the existing literature ($\eta_D \in 0.81, 1.5, 2.8$). The results of such calculations are all summarized in Table 2.2. In all cases, supply must be more elastic than demand, and the difference between the two must be substantially larger under low mark-ups, although far

[11] The value one for the *conduct parameter* results from the profit-maximizing markup under monopoly, which is the inverse of the elasticity of demand.

from constant marginal costs (perfectly elastic supply). Indeed, the second set of columns, which take the Lerner Index value estimated by O'Connell and Smith (2020) for the UK's soft drinks market, show that a relatively low elasticity of supply, just above the elasticity of demand, would be enough to result in full tax pass-through in most cases ($\eta_D \in 0.81, 1.5$). The partial derivatives of the model (Equations 2.6 to 2.8) show how variation on each of the three variables would affect tax pass-through. Larger pass-through ought to be observed in more concentrated markets and products with more inelastic consumers or more elastic supply.

Table 2.2: Calibrations of the model for full pass-through ($\rho = 1$)

Lerner Index = 0.16		Lerner Index = 0.46	
Demand Elasticity (η_D)	Supply Elasticity ($\hat{\eta_S}$)	Demand Elasticity (η_D)	Supply Elasticity ($\hat{\eta_S}$)
0.81	4.25	0.81	0.95
1.5	7.88	1.5	1.76
2.8	14.7	2.8	14.24

Notes: For every pair of the market concentration and demand elasticity parameters, the supply elasticity level that would yield full pass-through was calculated following Equation 2.5. For each combination of parameters in the table, any level of supply elasticity above the one calibrated here would yield overshifting ($\rho > 1$), and any level of supply elasticity below the calculated value would yield partial pass-through ($\rho < 1$). The conduct parameter (θ) in Equation 2.5 results from multiplying the Lerner Index times the elasticity of demand (η_D).

$$\frac{d\rho}{d\theta} = \frac{\frac{1}{\eta_D} + \frac{1}{\eta_S}}{\left(1 + \frac{\eta_D - \theta}{\eta_S} - \frac{\theta}{\eta_D}\right)^2} \quad > 0 \tag{2.6}$$

$$\frac{d\rho}{d\eta_D} = -\frac{1 - \frac{d\theta}{d\eta_D}}{\eta_S\left(1 + \frac{\eta_D - \frac{d\theta}{d\eta_D}}{\eta_S} - \frac{\frac{d\theta}{d\eta_D}}{\eta_D}\right)^2} \quad < 0 \quad : \quad \frac{d\theta}{d\eta_D} = \frac{p - mc - \tau}{p} \in (0, 1) \tag{2.7}$$

$$\frac{\partial\rho}{\partial\eta_S} = \frac{\eta_D - \theta}{{\eta_S}^2\left(1 + \frac{\eta_D - \theta}{\eta_S} - \frac{\theta}{\eta_D}\right)^2} \quad > 0 \quad : \quad \theta = \frac{p - mc - \tau}{p}\eta_D \in (0, \eta_D) \tag{2.8}$$

2.3.3 General equilibrium considerations

As summarised by Allcott et al. (2019a), several studies have reported a non-negligible degree of substitution between sugar-sweetened beverages and other drinks and foods. In the presence of non-zero cross-price elasticities, treating the taxed goods as an independent market can lead to a biased analysis of tax

incidence, even when the taxed goods only make a very small fraction of total expenditure. In those cases, partial equilibrium analysis may not be a good approximation to the general equilibrium tax incidence.

Substitution towards other goods can distort the prices of products not affected by the tax. For example, Dharmasena et al. (2014) apply an Equilibrium Displacement Model (EDM) to the soft drinks market, with multiple substitute goods facing different levels of tax. If supply is imperfectly elastic and the tax only affects one of the categories, substitution towards untaxed goods can effectively imply the demand curve for taxed goods will be steeper than partial equilibrium models would predict. That would result in a larger pass-through, and prices of the untaxed substitute good may increase due to a shift of their demand curve. Moreover, Agrawal and Hoyt (2019) present a simple alternative model that allows for overshifting even in a perfectly competitive market with perfectly elastic supply if the tax affects multiple inter-dependent markets and some of the affected goods have a price-elasticity that is smaller than their cross-price elasticity with other commodities.

The SDIL implemented in the UK only affects SSB with sugar content of at least 5 grams per 100 millilitres. The response of a large fraction of manufacturers by reducing the sugar content of their products resulted in very few brands on the higher tier and most of their substitute drinks remaining untaxed. Therefore, according to general equilibrium theory, a sugar tax like the one chosen by British legislators is likely to exhibit larger pass-through than other settings of soda taxes where all soft drinks are affected (including those with lower or null sugar content) and may have a spillover effect on prices of untaxed direct substitutes.

2.3.4 Elasticity of demand for online groceries

In the same way that consumers' purchasing decisions differ between on-the-go consumption and consumption at home, consumers may not behave identically when buying groceries online and when shopping at grocery stores in person. Moreover, different types of buyers may self-select into one channel or the other.

According to the ONS[12], the age group with the largest percentage of people purchasing food and groceries online is between 25 and 44, compared to those with ages above and below. There is mixed evidence on the heterogeneity of demand price elasticities by age groups. Estimates from smoking behaviour have pointed at young people as the most responsive group (Chaloupka and Wechsler,

[12]Office for National Statistics (2019).

1997; Franz, 2008), while other studies specific to SSB have found that the relation between demand elasticity and age follows a non-monotonic U-shape, with both younger and elder people being the most responsive to price changes compared to middle-aged groups (Muhammad et al., 2019; Dubois et al., 2020). In addition, selection also seems to occur by income, with the wealthier households being more likely to purchase their groceries online.[13] The existing studies looking at income group differences in the UK show that the lower-income groups are the most responsive to price changes of soft drinks (Briggs et al., 2013; Allcott et al., 2019b; Dubois et al., 2020). Therefore, internet shopping channels seem to attract more frequently the population groups with lower price elasticities overall.

Furthermore, consumption of goods purchased online is delayed in time and delivered to an address previously selected. This differs significantly from on-the-go purchases, which have been found to be more price-sensitive than consumption at home (Andreyeva et al., 2010; Powell et al., 2013; Benoit et al., 2016). Moreover, online purchases have a larger time delay between order completion and access to the purchased goods compared to shopping at physical stores, which can also affect consumer behaviour. Milkman et al. (2010) exploited exogenous variation in delivery times of online grocery purchases to estimate that people are more likely to buy healthier food when the time delay between order and delivery increases by several days. Therefore, if delayed consumption shifts the focus towards non-price attributes of the product, it can potentially reduce responsiveness to price changes.

On the other hand, online markets undoubtedly reduce search costs since access to information and comparison between sellers and products can be done from a single device. Price tracker tools even allow consumers to set alarms when the price of a selected good drops. This implies a reduction in the cost of price information. By making it easier for the buyer to look for cheaper alternatives, demand elasticity may increase (Stigler, 1961). However, better product information can also be accessed at a lower cost, shifting the attention of the buyer from the price to other product attributes and therefore decreasing elasticity (Anderson, 1968). Product information costs can play a relevant role in differentiated products with complex features, but that is less likely to be the case of sugary drinks that mainly differ in flavour, colour, size, ingredients and caloric content, all of which is printed as a label on the container itself.

Finally, customer loyalty also affects price sensitivity. Previous research found

[13] Dutton et al. (2013) found that respondents with yearly income above £30,000 were nearly three times more likely to have purchased groceries online within the last 12 months than those with yearly income below £12,500.

some evidence that consumers express more customer loyalty online than offline (Shankar et al., 2003; Danaher et al., 2003). Since the ability of consumers to choose between different alternatives is reduced when the range of alternatives they consider is smaller, higher customer loyalty could reduce the price elasticity of demand.

Therefore, from a theoretical perspective, it is not completely clear whether we should expect a larger or lower demand elasticity in online markets than in traditional brick-and-mortar stores. However, the existing literature specific to grocery shopping seems to consistently identify lower demand elasticity for online purchases than offline, and differences are strongest for food products (Degeratu et al., 2000; Chu et al., 2010).

2.3.5 Supply and competition in online markets

Similarly, firms may behave differently and face different cost structures when operating online versus offline. First, updating prices on websites is likely to be less costly than printing new labels and replacing the old ones physically in a brick-and-mortar store, which can explain a more immediate pass-through, as observed by Gorodnichenko and Talavera (2017). Second, the logistics of online supply do not require a customer-friendly display shelf with a large variety of products in a limited space, since orders can be fulfilled directly from a big warehouse. Therefore, adjustments to demand shocks in online sales mainly depend on stock capacity rather than display capacity. Those two factors can lead to a lower slope of the marginal cost curve in online grocery markets and corresponding larger elasticity of supply.

Another relevant factor affecting pass-through is the number of sellers in the market. The soft-drinks sector in the UK is highly concentrated at the producer level, with two big brands (Coca-Cola and PepsiCo) taking more than half of the market.[14] At the retailer level, the four biggest supermarket chains (Tesco, Sainsbury's, Asda, and Morrisons) gather 70% of all grocery sales in the UK.[15] The *big four* offer online grocery shopping services, where they compete against only two of their market followers (Waitrose and Iceland) and two purely online sellers (Amazon and Ocado). Aldi, Lidl, The Cooperative, smaller chains, and the corner shops did not offer any online grocery services by 2018. Moreover,

[14] According to data from IRI/Mintel, Coca-Cola and PepsiCo had a market share of 38% and 17%, respectively, in 2017-2018. Moreover, Coca-Cola Enterprises also owns Fanta and Schweppes, which increase its market share by an additional 9%.

[15] According to data from Kantar Worldpanel for 2017-2018.

Rafiq and Fulford (2005) analysed loyalty transfer from offline to online grocery shopping in the UK. They observed not only high retention but also that the market leaders attract a large proportion of their consumers from other market chains, further increasing market concentration. Big players may also benefit from technological entry barriers, since smaller businesses are unlikely to reach the top positions on search engines' results pages.

Last, intermediation can also affect pass-through since it generates a double-marginalisation problem. Given that the tax is levied on manufacturers, when those do not sell directly to final consumers but rather sell through independent retailers in an imperfectly competitive market, a double mark-up adjustment is expected to reduce pass-through (Hong and Li, 2017). This could explain the larger pass-through observed in private labels by Berardi et al. (2016) in France. The same way that retailers selling their own brands (most often through private labels) benefit from vertical integration, the internet offers a channel for producers to sell directly to consumers. Nevertheless, there is no evidence that online grocery sites selling a wide variety of independent brands systematically differ from traditional retail in that regard.

2.4 Empirical analysis

2.4.1 Description of the data

The dataset used in the analysis comprises a total of 134 items with daily prices covering between 6th April 2017 and 6th April 2019, a window of one year on either side of the date the SDIL came into effect, resulting in an unbalanced panel of 78,273 observations. Each observation includes the date, unique product identifier (ASIN[16]), product name, product brand, type of drink (category), type of container, container size, package (bundle) size, sugar content (grams per 100ml), price per litre, and corresponding SDIL tax per litre, which is calculated for non-exempt drinks based on their sugar content. Only products produced in the UK and sold and dispatched by Amazon are included in the analysis of this paper (third-party sellers and imported goods are excluded).

The dataset was created by the author using scraped information collected from Amazon UK using an *ad hoc* script in Python (van Rossum, 1995) and the price tracking tools Keepa and CamelCamelCamel. The items were identified by run-

[16]The Amazon Standard Identification Number (ASIN) is the unique identifier used by Amazon. It is equivalent to a Universal Product Code (UPC) but applies to each Amazon market separately (i.e. the same UPC would have a different ASIN in the UK and France).

ning searches on Amazon UK for the leading brand names of soft drinks and the keywords "milk", "tonic", "water", and "juice". While the product characteristics were scrapped directly from Amazon, the sugar contents data was collected from the official website of each manufacturer the week after the implementation of the SDIL (second week of April 2018) and again one year later to verify no changes. The categories used to classify each drink were: milk[17], still water, non-carbonated beverages (NCB)[18], cola-flavoured carbonated drinks, energy drinks, and other flavoured carbonated drinks. For the purpose of the analysis, drinks were further classified into different treatment groups based on their tax level, whether they were explicitly marketed as sugar-free, and their degree of substitutability with the drinks affected by the tax (Table 2.3). Overall, 20 items in the sample are liable to the tax, 21 items belong to the control group, and the other 93 items are untaxed soft drinks on which I explore potential general equilibrium effects (in the literature often referred to as *spillover*) given their role as close substitutes. The treated group has two levels of intensity, one for each band of the tax. The tax-free alternatives to the taxed drinks are also split into two groups: the *sugar-free* column includes beverages of brands that have a drink in the taxed categories (e.g. Coke, Pepsi, 7UP, Mountain Dew, or Red Bull) explicitly marketed as sugar-free (i.e. product names including Zero, Diet, or Free) ; the *other brands* column in Table 2.3 includes all soft drinks (carbonated and non-carbonated) of brands without any taxed products (e.g. Schweppes, Fanta, Irn-Bru, Rubicon, etc.). Many of the brands in the *competitor brands* group reformulated their drinks prior to the introduction of the tax, reducing the sugar content of their drinks below the SDIL threshold. The value of the tax was computed based on the date of the observation, sugar content at that time, and description of the drink. All prices and tax amounts are adjusted for inflation to April 2018 real value, and they are expressed in sterling pounds per litre.

An event study illustration in Figure 2.1 shows how the prices of taxed drinks and their closest alternatives (untaxed carbonated drinks) evolved during the year before and after the SDIL came into effect, compared to the evolution of prices of control drinks. The two depicted groups follow parallel trends with the control group before the event, even though seasonal effects may differ between drink categories. In particular, there seems to be an increase in prices of taxed drinks in summer and a drop in prices for carbonated soft drinks around the Christmas period in both years that is not observed in water and milk prices. Consistent with the theory of pass-through and general equilibrium effects, prices

[17]The category *milk* includes dairy-free alternatives such as soy milk.

[18]The category *non-carbonated beverages* includes all fruit juices, coco water, and iced tea drinks. Concentrated and powder formulas were excluded.

Table 2.3: Distribution of items across drink categories and treatment groups

Category	Group					
	Treated			Alternatives (tax-free)		
	Low SDIL	High SDIL	Controls	Sugar-free	Competitor brands	Total
Cola drinks	0	10	0	22	0	**32**
Energy drinks	0	4	0	3	0	**7**
Carbonated, other	3	3	0	5	36	**47**
Non-Carbonated, oth.	0	0	0	0	27	**27**
Milk	0	0	12	0	0	**12**
Water (still)	0	0	9	0	0	**9**
Total	**3**	**17**	**21**	**30**	**63**	**134**

Notes: An item is defined by an Amazon Standard Identification Number (ASIN), which is created the first time a new product is added to the Amazon catalogue. It is independent of who is selling it (Amazon or any third party) and allows tracking a single item across time and sellers.

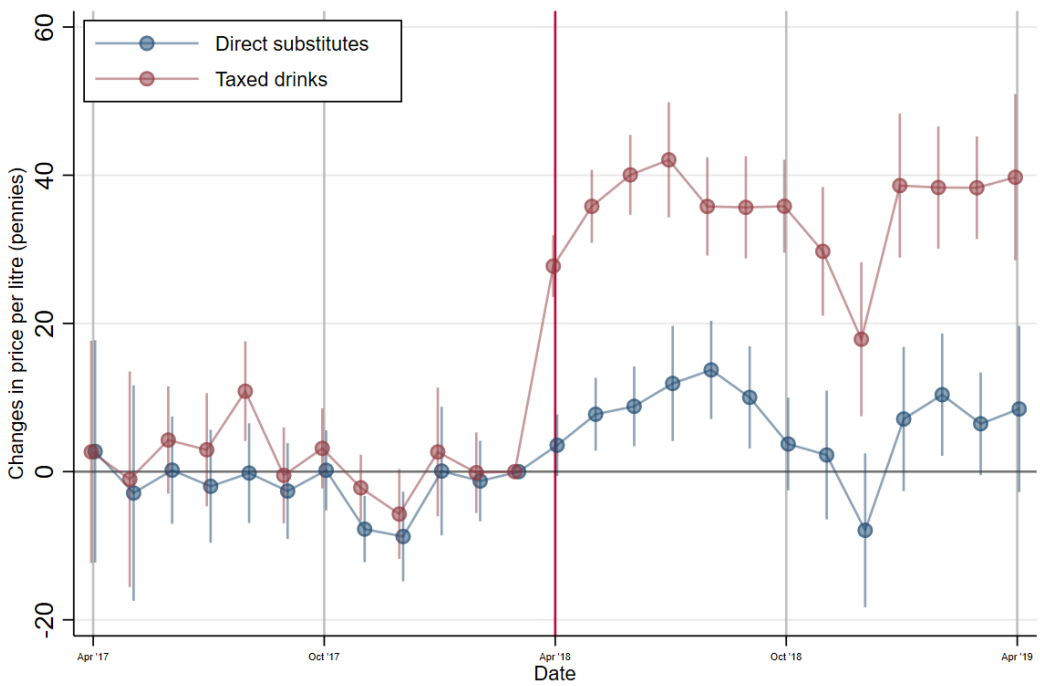
of taxed drinks surged immediately after the introduction of the tax, and prices of untaxed carbonated drinks (direct substitutes) also seem to increase after April 6th 2018, though to a much lower extent. Prices of drinks in the control group did not experience significant changes over the studied period.

2.4.2 Estimation method

To estimate the pass-through of the tax, I run a series of panel regressions with fixed effects at the item (i) level with unique ASIN code, fixed time (week) effects, and clustering errors at the brand level (e.g. Pepsi). Since prices of drinks seem to experience seasonality, I also control for category-season (cs) fixed effects. Treated items are those that belong to the categories subject to the tax and whose sugar content is above the threshold. The treatment variable is Tax_{it}, which takes value zero for all products before 6th April 2018, and the value of the tax per litre on affected drinks after that date. The control group is made of unflavoured milk and water products, exempted from the tax. Unflavoured water has been used as a control drink in previous evaluations of the sugar tax in Catalonia (Castelló and Casasnovas, 2020), France (Berardi et al., 2016; Etilé et al., 2018; Capacci et al., 2019), Mexico (Colchero et al., 2015) and Portugal (Gonçalves and Dos Santos, 2020), since it is the major ingredient of all soft drinks, is not affected by the tax, and the cross-price elasticity with SSB is low. However, Dubois et al. (2020) estimated a non-negligible level of cross-price elasticity between water and SSB in the UK. If water was indeed a substitute for SSB, using it as a control would bias my estimates of pass-through downwards. The same authors estimated a much lower cross-price elasticity for milk against all types of soft

drinks, which implies it is less likely to be affected by changes in the demand for soft drinks (with and without sugars). There are three other categories of drinks with larger cross-price elasticities with the SSB affected by the tax: sugar-free versions of the taxed products, other carbonated drinks of competitor brands with sugar contents below the tax threshold, and non-carbonated alternatives (NCB). I estimate whether their prices increased as a result of the SDIL even if they were not directly affected by it, testing the predictions of the general equilibrium models of pass-through (Dharmasena et al., 2014). The dependent variable (P_{it}) was calculated as price per litre, and the coefficients on the tax variable can be interpreted directly as the pass-through parameter (ρ). The parameter λ controls for any general shift of prices after April 2018 (the dummy $Post_t$ takes value one from 8th April 2018, for all drinks). All values are adjusted for inflation.

Figure 2.1: Monthly price changes by treatment group



Notes: The graph displays the differences of each group with respect to the evolution of prices of the control group (waters, milks, and dairy-free alternatives to milk). The connected dots in the graph correspond to the coefficients from a regression of nominal prices per litre (rescaled per item to their distance from the item's average price in March 2018) on a set of monthly dummies interacted with each group dummy, without controlling for any other covariates. Vertical spikes denote the 95 per cent confidence intervals. The group *direct substitutes* (blue) includes sugar-free alternatives as well as sugar-sweetened alternatives with sugar content below 5g/100ml and therefore exempted from the tax. The treatment group (red) includes drinks affected by either tier of the SDIL. The vertical red line denotes April 2018, the month when the SDIL came into effect. The coefficients for the month before that line show parallel trends of both groups with respect to the control group, with a seasonal peak in taxed drinks in August and a drop in prices of taxed drinks and direct substitutes in Christmas.

The first specification identifies the overall degree of tax pass-through on treated units by comparing treatment and control groups. Equation 2.9 estimates the target parameter ρ through differences in differences with item fixed effects, comparing the prices of taxed products and control products the year before the tax with the year after, on average. As previously mentioned, the affected drinks are liable at the point of first sale after production, and sellers could have stockpiled tax-free amounts before the introduction of the tax and consequently delayed the increase in prices. This could bias my estimate downwards. To test whether producers increased prices immediately or with some delay, I run a second specification as in Equation 2.10, which includes interactions with dummies for each quarter following the introduction of the tax (Q_1-Q_4). The interactions identify how long Amazon took to update prices. α_i represents item i's fixed effects and t is the time variable (day). Average pass-through at each quarter in the second specification is captured by the parameter ρ_q, where $q \in [1, 4]$ identifies quarters one to four following the introduction of the tax. In both regressions, I control for quarter seasonality within each category (cs).

$$P_{it} = \lambda_0 Post_t + \rho_0 Tax_{it} + \alpha_i + \xi_{cs} + \epsilon_{it} \tag{2.9}$$

$$P_{it} = \sum_{q=1}^{4} \lambda_{qt} Post_t \times Q_{qt} + \sum_{q=1}^{4} \rho_q Tax_{it} \times Q_{qt} + \alpha_i + \xi_{cs} + \epsilon_{it} \tag{2.10}$$

I then proceed to analyse the spillover effects of the tax on prices of the potential substitute categories. Similar to the first set of regressions, I first estimate the average spillover over the year after implementation (Equation 2.11) and run a second specification allowing for a delayed response by adding quarter dummies interacted with the spillover variable (Equation 2.12). To give an intuitive interpretation to the coefficients for spillovers so that they represent the price change on untaxed beverages in proportion to the tax on sugary drinks, the variable Z_{it} takes value zero before 6th April 2018 and the higher tier tax value after that date[19] only for items in any of the three substitute categories: sugar-free versions of the taxed products (SFV), untaxed carbonated beverages of products that do not have a taxed version (UCB), and untaxed non-carbonated beverages (NCB), as described at the beginning of this section.

[19]The value 0.288 corresponds to the tax per litre on a drink in the higher tier (with more sugar than 8g/100ml) including the proportional VAT. As with the price and the tax variables, the value of Z_{it} is adjusted for monthly inflation.

$$P_{it} = \lambda_0 Post_t + \rho_0 Tax_{it} + \pi_0 Z_{it}^{SFV} + \beta_0 Z_{it}^{UCB} + \kappa_0 Z_{it}^{NCB} + \alpha_i + \xi_{cs} + \epsilon_{it} \quad (2.11)$$

$$P_{it} = \sum_{q=1}^{4} \lambda_{qt} Post_t \times Q_{qt} + \sum_{q=1}^{4} \rho_q Tax_{it} \times Q_{qt} + \sum_{q=1}^{4} \pi_q Z_{it}^{SFV} \times Q_{qt} + \sum_{q=1}^{4} \beta_q Z_{it}^{UCB} \times Q_{qt} + \sum_{q=1}^{4} \kappa_q Z_{it}^{NCB} \times Q_{qt} + \alpha_i + \xi_{cs} + \epsilon_{it} \quad (2.12)$$

2.4.3 Heterogeneity across categories and package sizes

The products included in the sample are likely to have different demand elasticities (η_D) and face different levels of competition (θ). Given that all sugar-sweetened beverages (SSB) have a similar composition and logistical chain, there is no reason to think supply elasticity (η_S) should significantly differ across different drinks. I exploit differences in market competition and demand elasticity to test the model predictions explained in the theory section. For that purpose, I compare drinks liable to the tax with drinks in the control group, interacting the tax variable with dummy variables to identify heterogeneous tax pass-through on two dimensions: type of SSB (cola drinks, energy drinks, and other flavoured carbonated drinks) and container sizes (*small*, containing less than 360ml, *medium*, with between 360ml-999ml, and *big*, with one-litre capacity or more). Heterogeneity across drinks types has been previously studied in other settings where the tax affected a broader set of beverages. For instance, Falbe et al. (2015) found larger pass-through on sodas and lower on sweetened teas compared to non-pure fruit juices in Berkeley, and a good amount of papers have estimated higher tax pass-through for carbonated than for non-carbonated beverages (Colchero et al., 2015; Berardi et al., 2016; Silver et al., 2017; Capacci et al., 2019). However, carbonated SSB was the only category of drinks finally affected by the tax in the UK after other beverages were reformulated to reduce their sugar content below 5g/100ml[20], and evidence on heterogeneity of pass-through within this category is scarce, mainly comparing energy drinks to all other carbonated drinks (Campos-Vázquez and Medina-Cortina, 2019; Cawley et al., 2020). Given the leading role of cola drinks in the soft drinks market, it seems relevant to analyse whether this

[20]See Table B.2 in Appendix B.1 for a comparison of sugar contents of some of the most popular soft drinks in the UK, before and after the SDIL came into effect.

subset of sugar-sweetened beverages exhibits a different passthrough than other-flavoured carbonated SSB with smaller market shares. Regarding heterogeneity across sizes, several papers estimate over-shifting for single-serving containers (Rojas and Wang, 2017; Cawley et al., 2020). Other studies compare sizes of items sold in bundles but only account for two size groups: individual sizes up to one litre and family sizes above one litre. Colchero et al. (2015) and Powell et al. (2020) estimated larger pass-through for smaller (individual) sizes, consistent with the findings on single-serving containers. Nevertheless, the regular single-serving can of SSB usually contains 330ml (355ml in the USA), and there is no obvious reason to believe drinks between that size and up to one litre are consumed following a similar pattern. Dividing package sizes into three groups, with a middle-sized category, relaxes this assumption and allows me to test for a non-monotonic relation between container size and tax pass-through.

Table 2.4: Market competition and demand elasticities, by category

	Cola drinks	**Energy drinks**	**Other flavoured**
Lerner Index	$[0.51, 0.56]$	$[0.34, 0.48]$	$[0.53, 0.68]$
Elasticities of Demand:			
330ml-380ml cans	$[2.8, 2.99]$	2.72	$[2.94, 3.25]$
500ml bottles	$[2.36, 2.67]$	2.58	$[2.54, 2.69]$
2l bottles	$[1.44, 2.39]$	-	2.4

Notes: This table summarises the range between the lowest and the highest demand elasticities and mark-ups of sugar-sweetened carbonated soft drinks in each category and unit size estimated by Dubois et al. (2020) and O'Connell and Smith (2020) using data from the United Kingdom. Can sizes are 330ml for *cola drinks* and *other flavoured*, and 380ml in the case of *energy drinks*. The brands covered by the cited studies were: Coke, Cherry Coke, and Pepsi among *cola drinks*, Lucozade and Red Bull for *energy drinks*, and Dr Pepper, Fanta, Sprite, and Irn-Bru for the *other flavoured* group.

According to elasticity estimates for soft drinks in the UK from Dubois et al. (2020) and O'Connell and Smith (2020), other-flavoured sodas seem to face a slightly more elastic demand than cola and energy drinks (Table 2.4). Comparing across sizes, demand elasticity for small cans is larger than that of medium-sized bottles, while demand is the least elastic for big family-size bottles. Differences between package sizes are larger than those observed between beverage types. With regards to market competition, the two big producers in the UK, Coca-Cola Enterprises and Britvic, cover around half of all sales of soft drinks in the country.[21] In a report of the Competition Commission evaluating the merger of AG Barr and Britvic, gross margins for companies in the sector were estimated to be as large as 60 per cent (C. C. Inquiry Group, 2013). Using structural mod-

[21]Calculated using data from the *Soft drinks and non-alcoholic beverages in the UK* dossier from Statista, 2019.

elling, O'Connell and Smith (2020) estimated Lerner Index values[22] that reflect significantly lower mark-ups on energy drinks compared to the other two categories. According to the theoretical model discussed earlier in this paper,[23] the large market concentration of cola drinks is likely to lead to higher pass-through compared to energy drinks, with a much lower market concentration and similar elasticity of demand. Other-flavoured drinks, on the other hand, are the category with the largest mark-ups and more elastic demand, with each of these two variables affecting pass-through in opposing directions. That makes it difficult to predict whether other-flavoured drinks should exhibit larger pass-through than the other two categories or not. With regards to size heterogeneity, medium-sized packages should exhibit larger pass-through than smaller-sized items unless mark-ups were larger for the latter, and so should family-sized packages, contrary to the findings cited in the previous paragraph. Equation 2.13 describes the specification used to estimate differential pass-through across drink types and container sizes. As in the other regressions, price is expressed in sterling pounds per litre (P_{it}), the base drink category is *cola drinks*, and the variable Tax_{it} takes the value of the corresponding tax on each product from 6th April 2018. The dummies *Other* and *Energy* are interacted with the tax variable to capture the differential tax pass-through compared to cola drinks. The base container size is *medium*, and the dummies *Small* and *Big*, interacted with the tax variable, are used to estimate the additional tax pass-through on those sizes, compared to the medium size, on average, for all drinks.

$$\begin{aligned} P_{it} = \lambda Post_t + \rho_{cola} Tax_{it} + \rho_{other} \mathbb{1}\{Other_i\} Tax_{it} + \rho_{energy} \mathbb{1}\{Energy_i\} Tax_{it} \\ + \rho_{small} \mathbb{1}\{Small_i\} Tax_{it} + \rho_{big} \mathbb{1}\{Big\} Tax_{it} + \alpha_i + \delta_{cs_{it}} + \epsilon_{it} \end{aligned} \quad (2.13)$$

Table 2.5: Distribution of items across drink categories and container sizes

Category	**Container size**			**Total**
	Small (<360ml)	**Medium**	**Big (≥1L)**	
Cola drinks (taxed)	7	1	2	**10**
Energy drinks (taxed)	3	1	0	**4**
Carbonated, other (taxed)	3	1	2	**6**
Milk (control)	0	3	9	**12**
Water (control)	1	5	3	**9**
Carbonated, other (untaxed)	25	7	4	**36**
Total	**39**	**18**	**20**	**77**

[22]The Lerner Index is calculated as the ratio between the mark-up and the price of a product.
[23]Weyl and Fabinger (2013).

Table 2.5 shows the distribution of container sizes and drink categories in the treated and control groups. The last category, *Carbonated, other (untaxed)*, is not initially included in the analysis due to its potential role as a substitute (which could bias the estimates towards zero if general equilibrium effects were significant), but it is added to the control group in a second specification for robustness, increasing the sample size and thus improving the statistical power of the estimation. Container sizes are not evenly distributed across beverages, and thus the results from this regression should be taken with caution.

2.4.4 Results

In line with the simulation predictions of Dubois et al. (2020), the ex-post evaluation by O'Connell and Smith (2020), and the findings for most other countries that implemented the policy in large jurisdictions, my analysis shows that drinks affected by the SDIL in the UK experienced full pass-through of the tax to prices, on average. In addition, consistent with the predictions of a general equilibrium model with substitution towards untaxed goods, the price of sugar-free versions of the taxed drinks also increased, as would be expected if there was a shift of demand from the taxed sugary version to the untaxed sugar-free version. Table 2.6 shows the point estimates of tax pass-through on liable beverages and the spillover on substitute drinks due to general equilibrium effects. In columns 1 and 2, the coefficient for tax pass-through on affected drinks is at around 1.1, not statistically different from full pass-through. Furthermore, the specifications with quarter dummies (columns 3 and 4) show how prices were immediately updated within the first quarter after the introduction of the taxby its total amount (full pass-through). The coefficient on the fourth quarter is significantly higher than one at the 95 per cent confidence level, suggesting overshifting in the long run. Looking at the estimates of the spillover effect on untaxed substitutes (general equilibrium effects), only prices of direct sugar-free alternatives were affected, with an increase of nearly half the value of the higher tier tax. In this case, the increase in prices was smoothed across the first two quarters after the SDIL came into effect (column 4). No significant effect is found on prices of brands of carbonated beverages that reformulated their drinks before April 2018 or were already below the sugar threshold set by the SDIL, nor on non-carbonated alternatives. A graphical representation of the coefficient estimates is depicted in Figure 2.2.

Finally, the heterogeneity analysis reveals that the tax did not impact all drink categories and container sizes evenly (Table 2.7 and Figure 2.3). Cola drinks subject to the SDIL experienced overshifting of the tax on prices, estimated

Table 2.6: Tax pass-through of the SDIL and spillover on substitute drinks

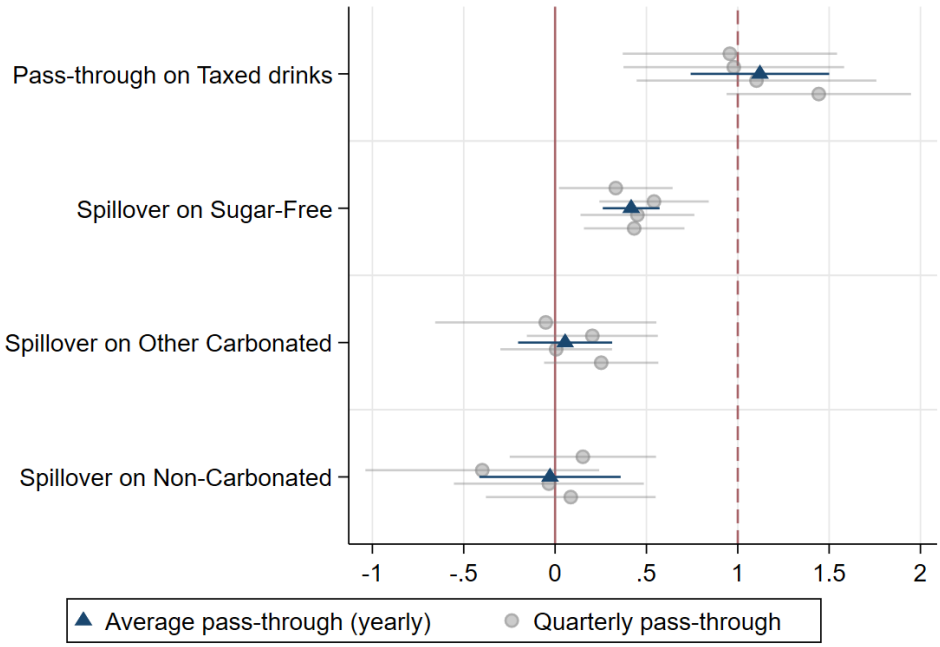
VARIABLES	(1) Average	(2) Average	VARIABLES	(3) Quarterly	(4) Quarterly
Tax Pass-through	1.115***	1.121***	Pass-through at Q1	0.979***	1.037***
	(0.190)	(0.186)		(0.311)	(0.289)
Spillover on Sugar-Free		0.417***	Pass-through at Q2	0.973***	1.090***
		(0.076)		(0.266)	(0.208)
Spillover on Other Carbonated		0.055	Pass-through at Q3	0.973***	0.943***
		(0.126)		(0.302)	(0.308)
Spillover on Non-Carbonated		-0.028	Pass-through at Q4	1.485***	1.397***
		(0.189)		(0.202)	(0.175)
			Spillover on Sugar-Free (Q1)		0.329**
					(0.149)
			Spillover on Sugar-Free (Q2)		0.521***
					(0.133)
			Spillover on Sugar-Free (Q3)		0.334**
					(0.137)
			Spillover on Sugar-Free (Q4)		0.512***
					(0.139)
Observations	23,831	78,273	Observations	23,831	78,273
R-squared	0.245	0.088	R-squared	0.253	0.094
Number of ASIN	41	134	Number of ASIN	41	134
Item FE	YES	YES	Item FE	YES	YES
Season-Category FE	YES	YES	Season-Category FE	YES	YES

Robust standard errors in parenthesis. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.
Notes: Columns 1 and 2 show the coefficients from regressions estimated following Equations 2.9 and 2.11, respectively. Columns 3 and 4 show the coefficients from regressions estimated following Equations 2.10 and 2.12, with quarter dummies, omitting the groups of drinks with coefficients not significantly different from zero. The full table with coefficient estimates for the impact of the SDIL on prices of substitute drinks, including the two groups omitted in this table (other untaxed carbonated drinks and other untaxed non-carbonated beverages), can be found in Appendix B.2 (Table B.3).

at 133% on average (column 1) and as large as 200% on medium-sized bottles (columns 3 and 4), which corresponds to an increase in prices of around 30%. As predicted by the partial equilibrium model of Weyl and Fabinger (2013), pass-through to prices of the category with larger demand elasticity, *other flavoured sodas*, was lower than to prices of cola drinks, with energy drinks also exhibiting a tax pass-through significantly lower than cola drinks, as expected from their considerably lower level of market concentration. Regarding heterogeneity across package sizes, the effect of size on pass-through was not monotonic, with medium sizes exhibiting the largest tax pass-through. The difference between small and medium packages seems to match the predictions of the model, based on the estimated elasticities (medium packages face a less elastic demand and hence experience a larger pass-through). However, larger containers (of 1 litre or more) seem to be the ones experiencing the lowest tax pass-through despite facing the least elastic demand. Albeit surprising, my results align with previous findings from Mexico (Colchero et al., 2015). Following the conclusions from the theoretical framework, a low pass-through on big sizes could only be explained

by very large mark-ups. Unfortunately, no other paper has estimated mark-ups separately for different container sizes to support or reject that hypothesis.

Figure 2.2: Average tax pass-through and spillovers of the SDIL



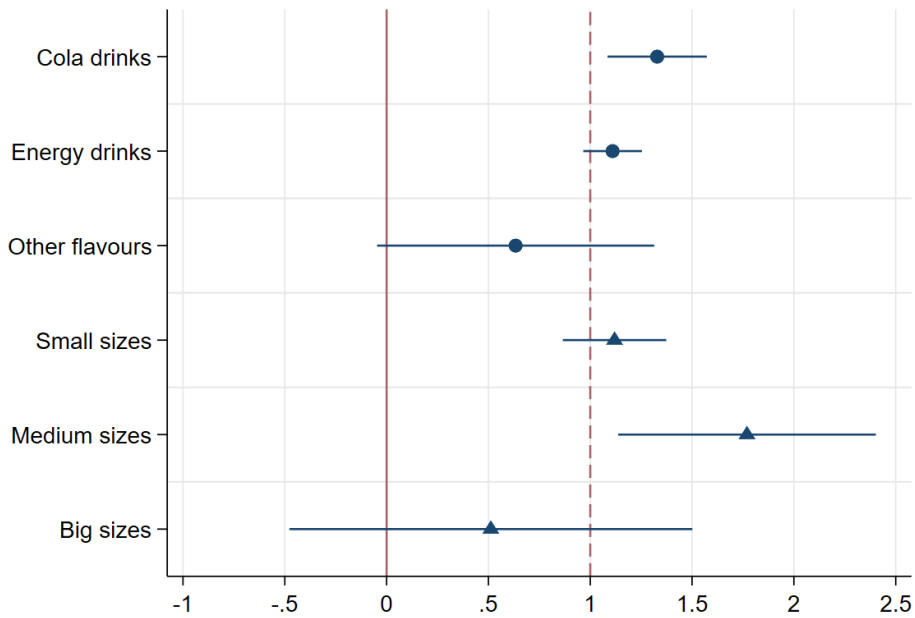
Notes: The graph displays the coefficients from column 2 in Table 2.6 in blue (triangles). Spikes represent 95 per cent confidence intervals. The grey markers correspond to similar regressions using only data from a given quarter of the year before and after the tax (e.g. April-June 2017 and April-June 2018). Ordered vertically, the first grey dot is the coefficient from a regression comparing prices of the first quarter after the tax with prices of the same quarter the year before. Each subsequent dot does the same with the second, third, and fourth quarters, respectively.

Table 2.7: Heterogeneous pass-through of the SDIL, by drink category and container size

VARIABLES	(1) Taxed vs Control	(2) Taxed vs Control	(3) Taxed vs Control	(4) Extended Control
Tax Pass-through	1.329***	1.770***	2.034***	2.025***
	(0.114)	(0.297)	(0.185)	(0.176)
Energy drinks	-0.219**		-0.434***	-0.439***
	(0.099)		(0.100)	(0.097)
Other-flavoured sodas	-0.695**		-0.578*	-0.560*
	(0.322)		(0.293)	(0.317)
Small container (<360ml)		-0.650**	-0.727***	-0.737***
		(0.243)	(0.170)	(0.164)
Big container (>999ml)		-1.258***	-1.213***	-1.237***
		(0.261)	(0.216)	(0.200)
Observations	23,831	23,831	23,831	44,609
R-squared	0.261	0.272	0.284	0.133
Number of asin	41	41	41	77
Item FE	YES	YES	YES	YES
Season-Category FE	YES	YES	YES	YES

Robust standard errors in parenthesis. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1. *Notes:* columns 1 and 2 evaluate heterogeneity in drink category and size separately, while column 3 evaluates both simultaneously, following Equation 2.13. The first three columns use milks and (still) waters as control drinks, while the last column also includes other (untaxed) carbonated from brands without taxed products as part of the control group. The inclusion of that group is only added for robustness, increasing the power of calculations, taking into account that such group of drinks was not found to exhibit general equilibrium effects and is therefore unlikely to bias the estimated coefficients (Table 2.6).

Figure 2.3: Average tax pass-through of the SDIL, by flavour and container sizes



Notes: The graph displays the coefficients from columns 1 (round markers) and 2 (triangular markers) in Table 2.7, linearly combined to represent the tax pass-through exhibited by every subgroup. The horizontal spikes represent the 95 per cent confidence interval.

2.5 Conclusion

This paper uses a unique dataset with daily prices of soft drinks sold by Amazon UK, a purely online retailer of increasing market share that has not been included in previous evaluations of the SDIL. By estimating a series of panel regressions, I find that SDIL taxes were fully passed through to prices of sugary drinks in the United Kingdom immediately after the introduction of the tax. General equilibrium effects also impacted prices of substitute drinks, with a spillover of nearly half the amount of the tax on untaxed sugar-free versions of the taxed drinks. While the change in prices of liable drinks occurred within the first quarter following the implementation of the SDIL, the spillover on untaxed sugar-free direct alternatives was spread over the two first quarters. Such significant spillover could indicate a demand shift from sugary to sugar-free versions within taxed brands and highlights the importance of general equilibrium effects when analysing soda taxes which are not levied on the entire portfolio of soft drinks. This finding is also relevant when interpreting the coefficients of other papers using diet drinks as a control (Colchero et al., 2015; Castelló and Casasnovas, 2020), implying those may be biased downwards. On the other hand, the lack of statistically significant spillover on other untaxed drinks implies that coefficients estimated using other alternatives such as fruit juices as a control group may not be significantly biased.

Furthermore, the results from the analysis on heterogeneity across products facing different levels of market concentration and demand elasticities qualitatively validate the predictions of a partial equilibrium model with differentiated products in an imperfectly competitive market. As predicted by the model presented by Weyl and Fabinger (2013), the category with the lowest demand elasticity and among the highest market concentration, cola-flavoured sodas, exhibits over-shifting of the tax (133%), with a pass-through estimate significantly larger than that of energy drinks and other flavoured sodas. The heterogeneity across container sizes is more surprising: the effect of size is not monotonic. Medium sizes have a significantly larger pass-through (with statistically significant over-shifting) than smaller or bigger sizes. The elasticity estimates for large containers in O'Connell and Smith (2020), lower than those of smaller containers, suggest that mark-up differences may be at the root of such differences.

Although the lack of consumption data restricts the analysis conducted in this study, this paper confirms the effectiveness of the SDIL in increasing prices of sugary drinks that were not reformulated to reduce their sugar content and provides evidence compatible with a demand shift towards their direct sugar-free alter-

natives. Moreover, it highlights the importance of accounting for heterogeneous pass-through across drinks when estimating optimal soda taxes or predicting expected tax revenues. The larger coefficient on cola-flavoured sodas, which exceeds one, reflects that sugar taxes in the United Kingdom significantly affected prices of the most popular category, inducing price increases of up to 30%, which is expected to lead to significant reductions in consumption of those beverages.

The main limitations of this paper are two-fold. First, my findings may not extrapolate to other countries discussing the implementation of sugar taxes that affect a broader range of products. The British Soft Drinks Industry Levy (SDIL) effectively triggered many producers of sugar-sweetened beverages to reformulate their drinks and reduce sugar contents below the SDIL threshold. This left very few products directly affected by the SDIL tax, which may affect substitution patterns differently than when a broader set of drinks is affected by a sugar tax. The second limitation of this study is the small sample. Although my panel of prices tracking uniquely identified SSB items over two years allows for a more robust specification than other studies using cross-sectional data, the subset of products included in the analysis is small, and it only covers one supermarket. Nevertheless, Amazon is the only supermarket chain that is present across a huge number of countries in several continents and is therefore of particular relevance given its potential expansion and market concentration. Moreover, it was one of the few supermarkets excluded in previous studies evaluating the SDIL. Overall, the findings of this paper are therefore an excellent complement to other studies in the UK using similar price data from other major supermarkets, such as O'Connell and Smith (2020) and Scarborough et al. (2020).

Finally, the findings of this study focusing solely on a purely online retailer do not significantly differ from other estimates of sugar tax pass-through using data from physical stores (Allcott et al., 2019a). This implies that calculations using online scraped prices of groceries may be a good approximation to price changes occurring in brick-and-mortar shops, reducing the costs of data collection.

3 Tax pass-through under psychological pricing: Lessons from the Spanish soda tax

3.1 Introduction

Corrective taxes on consumer goods have expanded in recent years from tobacco and alcohol to other unhealthy products like sugar-sweetened beverages (SSB).[1] The rationale behind the introduction of taxes on unhealthy food is that the consumption of such products imposes future costs on the consumers' health that they fail to take into account when making the purchase decision (internalities). Moreover, in countries with a welfare state, the consequences of consuming unhealthy food also generate costs to public health systems (externalities). The goal of corrective taxation as defined by Pigou (1920) is to shift the market equilibrium to the social optimum. In a perfectly competitive market, the tax will shift the supply curve upwards, hence generating a new equilibrium with a higher price and lower quantity. The exact equilibrium point and distribution of the burden between sellers and buyers will depend on the price elasticities of demand and supply. In imperfectly competitive markets, however, sellers with positive mark-ups on their products' prices may absorb the tax with no reduction in supplied quantities (Allcott et al., 2019a). In those cases, the tax would fail to induce a reduction in consumption.

The current evidence on pass-through and effectiveness of taxes on SSB is so far inconclusive. While Silver et al. (2017) found a large pass-through of the soda tax in Berkeley, leading to relevant reductions in sales, Rojas and Wang (2017) and Bollinger and Sexton (2018), among other authors, estimated much lower values for the exact same tax and city. Similar disparities between evaluations of a single tax can be observed in studies of sugar taxes in Chile (Caro et al., 2018; Cuadrado et al., 2020), France (Berardi et al., 2016; Etilé et al., 2018), Philadelphia (Cawley et al., 2018, 2020), or the United Kingdom (O'Connell and Smith, 2020; Scarborough et al., 2020). Most of these studies used different samples from the same country, either focusing on different types of stores or different beverage categories. Furthermore, Castelló and Casasnovas (2020) found very little change in consumption after the Spanish region of Catalonia introduced

[1] Before the year 2000, only Norway and Samoa had an excise tax on sweetened non-alcoholic beverages. In the second decade of the twenty-first century, over a dozen other big countries and several states in the USA introduced excise taxes on sugary beverages. Other countries like Denmark, Ethiopia or Samoa, also introduced taxes on saturated fats. For a summary of soda taxes implemented around the world, see Table B.1 in Appendix B.1.

a sugar tax with mandatory full pass-through by law, implying an elasticity of around 0.5. That value is significantly below the estimates of most existing studies analysing demand elasticities of sweetened beverages (Andreyeva et al., 2010; Powell et al., 2013). In this paper, I explore whether psychological pricing, a widespread strategy in grocery stores consisting of pricing items with *odd endings* (e.g. 0.99), can affect the level of tax pass-through observed after the introduction of taxes on sweetened beverages. Potentially, after a tax increase, psychological pricing could explain smaller than expected reductions in sales, since odd-ending prices are often underestimated and price changes within a given range may not be noticed by consumers. To test my hypothesis, I use the tax increase on sweetened soft drinks introduced in Spain in January 2021 to analyse the relation between tax pass-through and price-endings, as well as the frequency of psychological prices after the introduction of corrective taxes.

The behavioural economics branch of the literature on sticky prices has long acknowledged the role of psychological pricing techniques in generating price rigidity (Drakopoulos, 1992; Blinder et al., 1998; Herrmann et al., 2005; Knotek, 2011; Knotek et al., 2019). On one side, adaptation level theory argues that some values act as reference points, and consumers may only react to changes that go beyond an acceptance range around the reference point, often referred to as the latitude of acceptance (Monroe, 1971, 1973). On the other hand, there is evidence that left-digit processing (reading from left to right) leads to worse recall of the last digits of a number, resulting in an overestimation of prices with low numbers in the cents units and underestimation of prices with high numbers in the cents units. Both behavioural biases generate non-smooth demand curves with kinks at specific reference values. Increasing prices just beyond those thresholds can lead to a significant drop in marginal revenue since such a price increase will be accompanied by an unusually large reduction of demand (Blinder et al., 1998; Monroe, 2003). Therefore, the non-monotonicity of the marginal revenue curve as a result of kinked demand curves could explain the non-smooth distribution of prices of soft drinks and affect the pass-through of taxes on unhealthy drinks, impacting the effectiveness of such measures.

Using a panel of web-scraped daily prices of carbonated soft drinks, vegetable milks, water and alcoholic beers in three large Spanish supermarket chains, I exploit a value-added tax (VAT) code reform that doubled the taxes on sweetened soft drinks from 1st January 2021 in Spain. Before the introduction of the tax, 35 per cent of the non-alcoholic carbonated sweetened drinks in my sample had prices ending in either a nine or a zero, and psychological values in the first decimal were observed in around two of every three prices of that category of

beverages. That is above a 50% excess density of products at psychological prices compared to a uniform distribution across price endings. The proportion of items with psychological prices did not increase after the introduction of the tax, and the probability of an item to be priced with a psychological ending did not significantly depend on the products' initial price endings. Nevertheless, items with initial prices ending in zero had around a ten per cent lower probability of increasing than those with other endings. However, their average pass-through was not significantly lower than that of products with other endings, implying the lower probability to change was compensated by greater increases. Indeed, the pass-through of products initially priced at round integers (double-zero endings) significantly exceeded one. Products initially priced with endings .90 to .99 also had a five per cent lower probability of increase within the first month, but that difference disappeared in the second month, suggesting only a very short lag. Furthermore, another relevant psychological threshold affecting tax pass-through was the 50 cents ending. Products initially priced with endings in .40 to .49 exhibit a tax pass-through 30 percentage points lower than items with other endings.

The findings of this study contribute to a better understanding of the market responses to sugar taxes while being of relevance to the more general literature on sticky prices and psychological pricing. This is the first paper to evaluate the pass-through of the tax increase for sweetened drinks introduced in Spain in 2021. Sweetened drinks increased their prices by nearly the full amount of the tax within only two months from the day it came into effect. Furthermore, the probability that a treated product was priced with *odd endings* declined between December 2020 and April 2021 by more than that of control products, suggesting that sellers are not systematically recurring to such marketing technique to prevent consumers from reducing their consumption of unhealthy drinks in response to sugar taxes. Therefore, psychological pricing is unlikely to be the reason why some papers find very little price elasticity of consumption of sweetened drinks despite the full pass-through of sugar taxes (Castelló and Casasnovas, 2020). Finally, the finding of halves of Euro playing a role in significantly reducing tax pass-through implies that the impact of psychological prices on price stickiness is not limited to small cost changes like those studied from VAT reforms in Israel (Knotek et al., 2019) but exists even when costs increase by as much as 10%. Moreover, my results highlight the relevance of the first decimal digit of initial prices in determining tax pass-through to prices of soft drinks rather than the last digit alone. Therefore, future studies evaluating price rigidity should not limit their analysis to the effect of nine or zero endings but also explore the role of the

tenths of cents around the half and unit thresholds. Finally, the overshifting that I identify on products initially priced at round integers could be an explanation for the larger pass-through identified in on-the-go consumption from corner stores in settings where small stores are more likely to use round prices.

The rest of this paper is structured in five more sections. Sections II and III provide an overview of the literatures on psychological pricing and soda taxes, respectively. Section IV gives details of the data that I collected from the supermarket websites and presents some summary statistics. Section V includes the complete empirical analysis, starting with the average tax pass-through and separately estimating the role of odd endings at the last digit and the combined effect with the first decimal digit. To conclude, section VI highlights the study's limitations and the policy implications of the findings, and introduces the planned extensions of this project.

3.2 Psychological pricing and price rigidity

When shopping at any retail store, it is common to see a disproportionate number of price tags with odd endings (those right under round numbers). The practice of pricing items at values ending in nine can be traced back more than a hundred years ago and has been well documented in several surveys, particularly in low-priced goods (Hower, 1943; Twedt, 1965; Kreul, 1982). From a marketing perspective, it is believed that such endings are effective to maximise profits. In a series of experiments, Anderson and Simester (2003) found that prices ending in nine increased sales by up to 40 per cent compared to the closest five-ending prices right above and below. At the same time, they found no significant difference in sales between the two prices ending in five, despite their ten cents difference. A long list of studies[2] have also analysed the effects of nine-endings in different contexts and concluded that, for some categories of products, prices ending in nine result in greater demand than a smooth demand curve would predict. An exception to such a finding is the field experiment conducted by Bray and Harris (2006) in the UK. After rounding prices of products previously priced at nine-ending amounts in treated stores, sales of those items increased compared to the same products in the control stores. Nevertheless, the authors documented a long-existing practice at the retail chain of the field experiment to price all their non-promotional items at nine-ending values, which may have conditioned the way consumers evaluate prices in that particular store.

[2] A very comprehensive summary of the literature can be found in Table 1 of Lopez-Pastor et al. (2020).

The two main explanations for the widespread use of psychological prices are their image effect and the left-digit bias. The image effect refers to the fact that consumers relate odd endings to price deals and therefore assign a *low-price meaning* to nine-ending prices (Schindler, 2006). The left-digit bias, on the other hand, is caused by the higher attention that consumers pay to the leftmost digits compared to the rightmost digits of a number (Poltrock and Schwartz, 1984; Hinrichs et al., 1981). Thomas and Morwitz (2005) conducted a series of experiments to test whether left-digit bias is consistent with a model of numerical cognition where individuals read numbers from left to right and consequently compare numbers starting from the left, disregarding the rightmost digits after a difference is identified. The authors found evidence of left-to-right processing, with higher response latency when comparing numbers whose difference was found only at the rightmost digits. Their research also revealed that perceived distance between two numbers is driven by the leftmost digits, with odd endings magnifying perceived distances when the gap between the two numbers is small (e.g. comparing 2.99 to 3.00).

The focus on the left digits is believed to cause a worse recall of odd-ending prices. Schindler and Wiman (1989) conducted an experiment where participants were shown prices ending in 00 and other prices ending in 99 and 98. After asking participants to recall the prices for some of the items two days later, they found that the accuracy of their memory was worse for rightmost digits than for leftmost digits, therefore leading to underestimating prices with large odd endings (i.e. 99 and 98). In their experiment, even-ending (round) prices were recalled correctly twice as likely as odd-ending prices, and their probability of being underestimated was ten per cent lower. In another experimental study, Bizer and Schindler (2005) also provide evidence consistent with last digits drop-off. The researchers asked respondents to estimate how many products from a list could be purchased with 73 US dollars. Participants seeing prices ending in 99 estimated they could afford to buy significantly more products than those seeing comparable 00-ending prices. To quantify the bias, Strulov-Shlain (2019) used scanner data from the USA to structurally estimate that consumers react to a one cent change from a 99 ending to a round ending as if it were an increase of between 15 to 25 cents. Such amplified elasticity at round price points creates a region of dominated prices and could explain the bunching of products below psychological thresholds followed by a lack of products priced at low odd endings.

While the empirical evidence on amplified elasticities at psychological price points is mixed and the behavioural biases induced by odd endings do not affect all types

of products and sales channels equally,[3] industry practitioners seem to believe that demand reacts strongly around specific price points and they price their products accordingly, as with kinked demand curves (Blinder, 1991). Figure 3.1 graphically summarises the idea of kinked demand curves and their impact on marginal revenue in imperfectly competitive markets. Increasing the price from a value ending in 99 to the following round number triggers a large reduction in consumption, thus reducing marginal revenue despite the higher price per unit. When a product is priced at such odd-ending price points, it is possible that a small tax would not shift the marginal cost curve enough to induce a reduction in the quantity supplied, resulting in zero passthrough (Figure 3.2). This was the case analysed by Knotek et al. (2019), exploiting small VAT changes in Israel. Throughout the period covered in their study, the VAT suffered four small increases and four small decreases of between 0.5 and 1 per cent. Using a large sample of product prices from different convenience stores, the authors found that products initially priced at *favoured endings*[4] were significantly less likely to adjust their price after a change in VAT than products with other price endings. The difference in probability of adjustment is larger when only looking at price increases: after a positive change in VAT, products with favoured endings were only half as likely to increase compared to other price endings.

3.3 Soda taxes

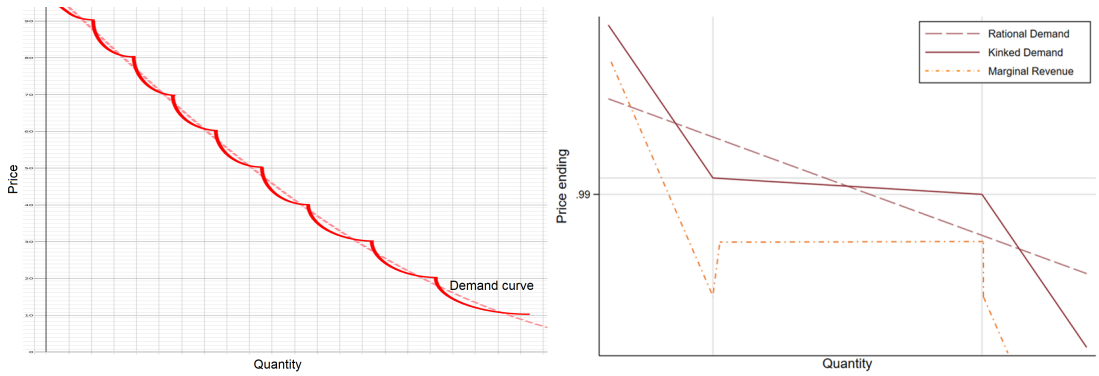
In recent years, a growing number of countries have introduced taxes on sugary drinks to promote healthier diets. The World Health Organisation (WHO, 2015a) published a report in which it recommended that taxes aimed at reducing consumption of unhealthy food and drinks should induce significant price changes (around 10 to 20 per cent minimum) to effectively change consumption habits. They issued recommendations on three types of taxes: specific excise taxes per volume or unit, *ad valorem* excise taxes, and value-added taxes.

Per-volume excise taxes have been mostly adopted by European countries (per litre) and North American States (per fluid ounce). Ad valorem excise taxes on sugary drinks, on the other hand, are more frequent among Latin American and Asian countries. Variation in value-added taxes (VAT) is often very restricted (most countries only have a regular rate and a reduced rate), which explains their rare use for corrective taxation purposes. Nevertheless, in 2021, Spain changed

[3] Pauwels et al. (2007); Lopez-Pastor et al. (2020).

[4] Knotek et al. (2019) use the concept *favoured ending* to refer to the price ending that a shop uses for more than fifty per cent of its products. They document that prices ending in zero are preferred by small convenience stores while nine-endings are preferred by large grocery stores.

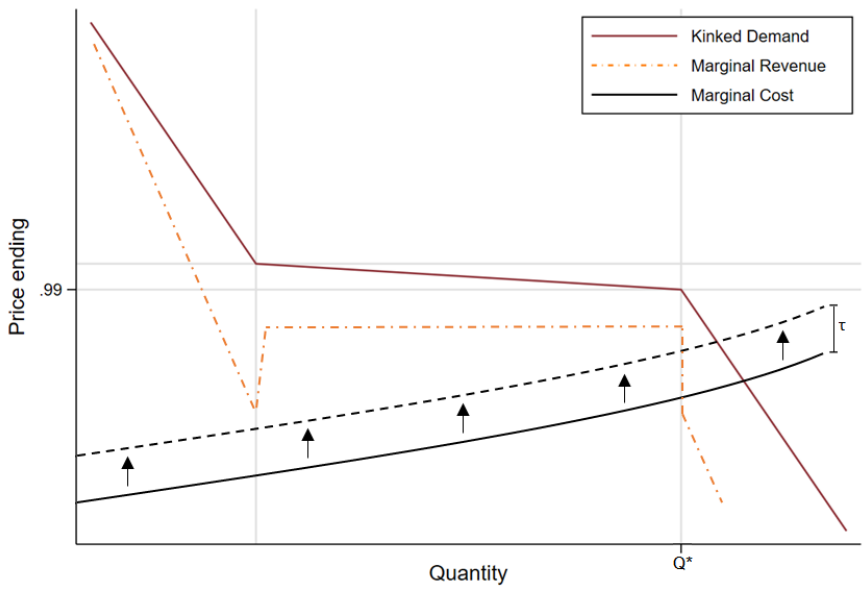
Figure 3.1: Kinked demand curves and marginal revenue



(a) Kinked demand curve (b) Marginal revenue at kinks

Notes: The kinked (behavioural) demand curve (solid) is superposed on a smooth (rational) demand curve (dashed) in both charts. Chart (a) represents a demand curve with several reference price points. Chart (b) focuses in a single price point, the round ending. Price changes that involve a change of the unit digit are most salient to consumers, amplifying the elasticity of demand around the round ending. Changes in prices that do not affect the digits at the left of the decimal point are less noticed by consumers, falling in what is often called the *acceptance regions*, with reduced elasticity of demand. Under monopolistic competition, the marginal revenue (yellow dot-dashed line) significantly increases when reducing the price right below the reference price point, since the small decrease in price is offset by a large increase in quantity.

Figure 3.2: Effect of a small tax in an imperfectly competitive market with a kinked demand curve



Notes: The shift of the marginal cost curve generated by the tax (τ) does not change the optimal quantity produced (Q*) and resulting optimal price, since the intersection between the marginal cost (black line) and the marginal revenue (yellow) curves remains at the same quantity level.

Table 3.1: Comparison of soda taxes in Europe and North America.

	33cl/12oz can of Coke		1l bottle of Coke			
Country/State	**Price**	**Sugar Tax**	**Price**	**Sugar Tax**	**Tax/litre (incl. VAT)**	**Supermarket**
Belgium (2016)	0.71 €	3%	1.75 €	4%	0.0722 €	Carrefour
Estonia (2018)	0.79 €	15%	1.19 €	30%	0.36 €	Selver
France (2012)	0.53 €	5%	1.23 €	6%	0.0755 €	Carrefour(2)
Hungary (2011)	225.00 Ft	3%	193.71 Ft	10%	18.75 Ft	Tesco(1)
Ireland (2018)	1.15 €	9%	1.60 €	19%	0.30 €	Tesco(1)
Portugal (2017)	0.72 €	9%	1.39 €	15%	0.20 €	Auchan
Spain, Catalonia (2017)	0.75 €	16%	1.10 €	22%	0.13 € + 10%	Carrefour
Spain (2021)	0.75 €	10%	1.10 €	10%	10%	Carrefour(1)
United Kingdom (2018)	£ 0.50	19%	£ 1.45	20%	£0.29	Tesco(2)
USA, Albany (2017)	1.43	9%	0.91	14%	0.37	Walmart(2)
USA, Berkeley (2015)	1.43	9%	0.91	14%	0.37	Walmart(2)
USA, Boulder (2017)	1.43	18%	0.91	28%	0.73	Walmart(2)
USA, Oakland (2017)	1.43	9%	0.91	14%	0.37	Walmart(2)
USA, Philadelphia (2017)	1.43	13%	0.91	21%	0.54	Walmart(2)
USA, S. Francisco (2018)	1.43	9%	0.91	14%	0.37	Walmart(2)
USA, Washington (2010)	1.43	16%	0.91	25%	0.64	Walmart(2)
Mexico (2014)	14.00	8%	21.00	8%	1.00 + 8%	Walmart

Source: Elaborated by the author. Coke prices were extracted from each supermarket website in April 2021 and are merely used as a reference for cross-country comparisons.
Notes: (1) In the absence of 1 litre bottles, I display the per litre price of the first available bottle size above 1 litre. (2) In the absence of single cans, I display the price per can of the smallest pack available.

the classification of sweetened drinks, moving them from the reduced category (the one that applies to all food and non-alcoholic beverages) to the regular band, effectively imposing a 10% *soda tax* through that mean.

Despite the recommendations of the WHO, the taxes introduced in most countries are not too large. Table 3.1 summarises the taxes implemented in some European and American countries and states. Many of the tax amounts displayed in the table are already higher than their initial values when they were first approved, and yet non of the examples presents a tax that is above 20% of the price of a regular can of coke (33cl size in Europe and 12oz size in America).

Changes in prices as a result of those taxes have rarely achieved the recommended increase of 20%. As presented in the previous section, the combination of a low tax and a high concentration of products priced at psychological thresholds could explain the low tax pass-through estimated in some countries where other explanations like cross-border shopping are not plausible. Etilé et al. (2018) found a low pass-through (39%) of the sugar tax introduced in France in 2012 using home-scan purchase data, although Berardi et al. (2016) and Capacci et al. (2019) found full pass-through evaluating the same French policy. Caro et al. (2018) estimated partial (50%) pass-through of the sugar tax on carbonated SSB in Chile, while Cuadrado et al. (2020) estimated overshifting using a different dataset for the same country. In the United Kingdom, Scarborough et al. (2020) estimated partial (50%) pass-through of the Soft Industry Levy Tax, while O'Connell and Smith (2020) estimated full pass-through, in line with my results included in

Chapter II of this volume, each of the three studies using different sets of web-scraped prices. The reasons driving these differences are not obvious, and they are unlikely to be purely the result of differences in elasticities of demand and supply or the level of market concentration, given that differences also appear between studies within the same markets, when using different sources of data.

Psychological prices have been previously linked to price rigidity, yet their role in determining the pass-through of soda taxes remains unexplored. Using the Spanish reform and prices from three big supermarket chains, I evaluate the relation between price endings and soda taxes. On January 1st 2021, Spain changed its value-added tax (VAT) law to exclude sweetened drinks from the reduction in VAT that applies to all nutritional products, hence imposing an additional 10% tax on all non-alcoholic drinks with added sugar or other sweeteners.[5] While volume-based taxes affect items differently based on their price per litre, which is usually decreasing with package size, proportional taxes like the value-added tax place the same burden on all products. Therefore, the policy reform introduced in Spain affects all packages equally, independently of unit sizes.

3.4 Data

On Wednesday 2nd of December, 2020, the Spanish Congress approved the reform of the VAT Law that moved non-alcoholic drinks with added sugar or sweeteners from the reduced category (10% VAT) to the regular category (21% VAT), resulting in a *de facto* soda tax of 10% on previous prices. That same week, I started scraping prices from the websites of three major supermarkets (Alcampo, Eroski, and Mercadona) using an *ad hoc* script programmed in Python (van Rossum, 1995). Following products in the categories of energy drinks, sports (isotonic) drinks, tonic sodas, and cola, lemon and orange flavoured carbonated sodas, I built a panel dataset covering daily prices of 501 taxed items from 101 brands, from four weeks before the tax change until four months after it. I also followed all mineral still and sparkling (unsweetened) waters, alcoholic beers, and vegetable milks sold throughout the same period, which were unaffected by the tax change, adding 574 control products from 167 different brands (Table 3.2). In the following pages, I refer to drinks affected by the tax change as *treated* or *taxed* drinks, and to unaffected drinks as *control* or *untaxed* drinks.

The distribution of prices before the introduction of the new tax was not even.

[5]It is relevant to note that although the tax change was an increase of the VAT in 11 percentage points (from 10% to 21%), the actual tax is only charged on the VAT base, therefore equivalent to a 10% tax on the previous tax-included price, which is the value shown in price tags in Spain, by law.

Table 3.2: Summary of the data, by retailer and product category

	Cola sodas	Lemon sodas	Orange sodas	Energy drinks	Sport drinks	Tonic soda	Sparkling Water	Still Water	Veg. Milks	Alc. Beers	
Alcampo											**Total**
Price/litre (eur)	2.84	1.37	1.11	1.14	1.52	3.42	1.90	0.70		1.67	
Container size (ml.)	419	938	1068	1011	819	332	733	997		360	
Bundle size (u.)	2	4	2	2	2	2	2	2		5	
Number of items	37	94	44	29	41	47	32	98	0	68	490
Number of brands	7	7	11	6	8	11	12	23	0	16	101
Private Label items (%)	0%	29%	18%	33%	25%	9%	8%	4%		6%	13%
Eroski											**Total**
Price/litre (eur)	2.67	1.51	1.16	1.12	1.70	2.76	1.30	0.83	1.61	3.05	
Container size (ml.)	416	975	1020	1084	780	342	854	957	1000	359	
Bundle size (u.)	2	4	2	2	2	1	1	1	1	4	
Number of items	22	60	21	19	23	17	14	54	44	195	469
Number of brands	5	3	6	4	4	6	7	18	10	46	109
Private Label items (%)	20%	33%	17%	25%	25%	17%	14%	6%	10%	2%	11%
Mercadona											**Total**
Price/litre (eur)	1.84	1.03	0.77	0.45	0.96	1.60	1.11	0.50	1.36	2.70	
Container size (ml.)	563	1304	1499	2000	1357	512	958	1076	953	333	
Bundle size (u.)	3	4	2	1	1	6	4	6	5	6	
Number of items	8	12	10	4	7	6	6	13	18	32	116
Number of brands	5	3	5	2	4	4	4	8	3	20	58
Private Label items (%)	40%	33%	20%	50%	50%	25%	50%	13%	33%	10%	23%
All Supermarkets											**Total**
Number of items	67	166	75	52	71	70	52	165	62	295	1075
Number of brands	17	13	22	12	16	21	23	49	13	82	268
Private Label items (%)	11%	31%	18%	32%	27%	12%	15%	5%	17%	4%	13%

Notes: The first six groups are *treated* categories, while the four last groups are *control* categories. The table summarizes the average price per litre, average container size, average bundle (pack) size, and the total number of items and brands in each category and supermarket, as well as the percentage of items that are private label.

Endings in zero and nine represent more than a third (36%) of all December 2020 prices included in my dataset and more than two thirds (68%) of prices in specific categories and sellers (Table 3.3). The bunching of prices at such strategic endings seems to indicate that sellers in the Spanish market take into account psychological thresholds and left-digit bias. The frequency distribution of the last digit, however, is not as unbalanced as that observed by Ater and Gerlitz (2017) or Knotek et al. (2019) in Israel, where most stores had more than 50% of their product prices ending in one specific digit (either nine or zero).[6]

In the setting I analyse in Spain, however, given the magnitude of the tax change (+10%), the second last digit (first decimal) may also play an important role. Moreover, the distribution of both last digits may not be independent of each other. As depicted by Figure 3.3, zeros and nines in the last digit were most often preceded by the same number in December 2020. Meanwhile, the most frequent first decimal numbers overall were four and five, which are not usually classified as psychological thresholds but seem to be relevant in this case. Therefore, while most of the literature has focused on the role of the last digit only, I exploit the large shock to sweetened drinks the Spanish VAT reform to further explore their role in combination with different preceding numbers.

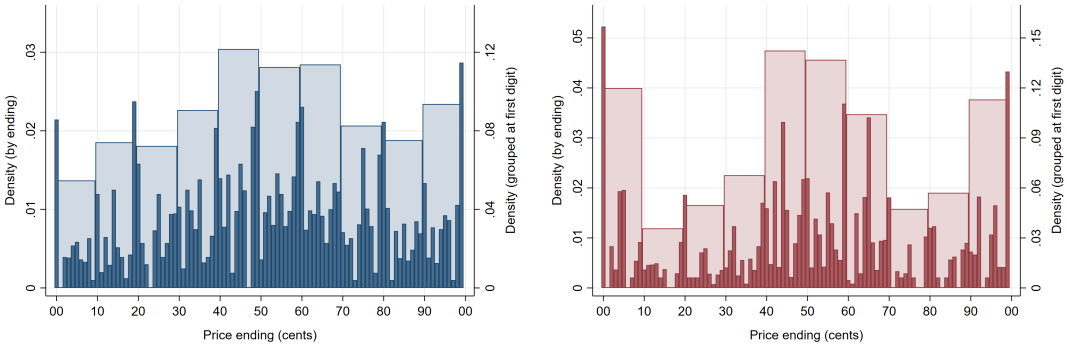
[6]Full detail of the distribution of price endings by drink category and seller can be found in Appendix C.2.

Table 3.3: Proportion of beverage prices ending in nine or zero in December 2020, by seller and category

	Cola sodas	Lemon sodas	Orange sodas	Energy drinks	Sport drinks	Tonic sodas	Sparkling Waters	Still Waters	Veg. Milks	Alc. Beers	Overall
Alcampo	40%	34%	36%	38%	44%	27%	40%	30%		21%	33%
Eroski	68%	29%	45%	43%	39%	58%	19%	28%	47%	36%	37%
Mercadona	25%	58%	60%	25%	57%	0%	50%	23%	78%	44%	47%
Overall	42%	34%	41%	38%	44%	28%	37%	29%	56%	32%	36%

Notes: This table summarises the proportion of prices in each category-supermarket pair ending in either a nine or a zero in December 2020, what I consider *psychological endings*. If price endings were uniformly distributed, the proportion of prices ending in a nine or a zero would be 20 per cent. However, for most categories in the three supermarket chains, the proportion was above 30 per cent, implying an excess mass of more than 50%. The last column shows the proportion of psychological endings among prices in all ten categories within each supermarket, while the bottom row shows the proportion of prices with psychological endings within each category, across the three supermarkets.

Figure 3.3: Distribution of endings of product prices in December 2020



(a) Price endings of control drinks (b) Price endings of treated drinks

Notes: These bar charts display the distribution of the endings (cents) of prices of untaxed (left) and taxed (right) drinks in December 2020, before the new tax on sodas was introduced. The thin bars depict the density of each price ending (two digits), while the wider bars (shaded in a more transparent colour tone) represent the total density added at each first digit block (tenths of cents). Final digits are not evenly distributed among first digits in either group, with nine-ending being most often preceded by another nine and zero-ending most often preceded by another zero.

3.5 Empirical analysis

To assess the role of psychological prices on tax pass-through, I classify products into price ending groups based on the ending of their most frequent price in December 2020, which I use as the *reference price* of a product before the tax change. I start my analysis by providing an overview of the evolution of prices of treated and control drinks in general, calculating the average pass-through of the new Spanish tax on sweetened drinks. After that, I proceed to evaluate the differential impact that psychological endings (zero and nine) had on tax pass-through and price rigidity. To complement the analysis, I also study such price endings in combination with different preceding numbers in the first decimal

position. I pay particular attention not only to the most natural threshold at the integer (round) full Euro, but also at the half-Euro threshold, where I observe most of the bunching of prices in December 2020. Note that the regressions estimating the effect of initial price endings assume that a treated item would evolve parallel to a control item with the same initial ending rather than to the average price change among control drinks. Finally, I further explore the price-setting strategy of supermarkets in Spain by evaluating the changes in the probability of prices having psychological endings in response to the new tax, which could be used to (mis)lead consumers into underestimating the changes.

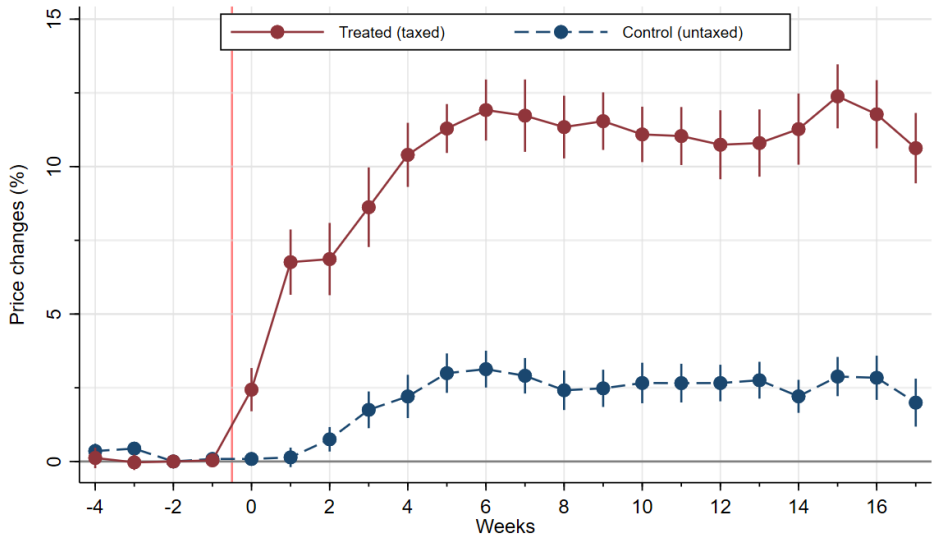
3.5.1 Average pass-through of the Spanish soda tax

Prices of treated and control drinks followed a parallel trend before the tax change (Figure 3.4). From the first week of January 2021, when the new VAT tax code came into force, treated drinks increased prices departing from the trend of control products. Both groups seem to recover parallel trends after around six weeks, suggesting that supermarkets passed the tax through to consumer prices in less than two months. To estimate the actual pass-through parameter (ρ), I run a panel regression following the specification in Equation 3.1, where the dummy T_i identifies drinks affected by the tax change and $Log(P_{it})$ is the logarithm of the price of product i on day t. December ($m = 0$) is the base category for the months variable (m) and α_i represents the item fixed effects. In some of the specifications, I also control for supermarket-month fixed effects X_{it}.

$$Log(P_{it}) = \beta_0 + \sum_{m=1}^{4} \beta_m \mathbb{1}\{Month_t = m\} + \lambda T_i + \sum_{m=1}^{4} \rho_m T_i \times \mathbb{1}\{Month_t = m\} + \Omega' \mathrm{X}_{it} + \alpha_i + \epsilon_{it} \quad (3.1)$$

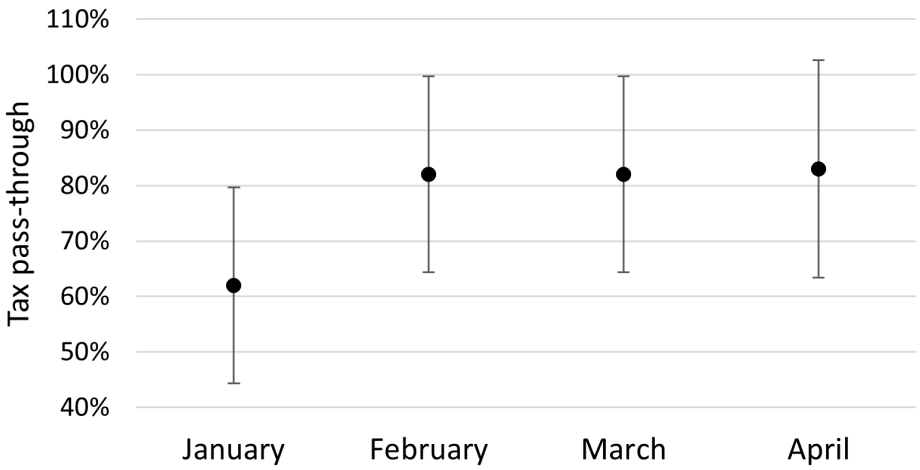
On average, supermarkets passed above 80 per cent of the tax to consumer prices. As shown in Figure 3.5, while pass-through was only partial in the first month (62%), the estimated coefficients are not significantly different from one (full pass-through) from February and remain stable throughout the following months (March and April). This implies that supermarkets updated their prices within the first month.

Figure 3.4: Evolution of prices before and after the tax change.



Notes: This graph plots the coefficients of a panel regression of the logarithm of prices on week dummies interacted with a dummy for each treatment group (taxed and untaxed), with item fixed effects. Vertical spikes represent 90% confidence intervals, with errors clustered at the brand level. Time zero represents the first week affected by the tax, and the vertical red line separates the before and after-tax change periods. Both groups seem to follow a parallel trend before the tax change, while prices of taxed drinks increase at much higher rates than the control group over the first six weeks after the introduction of the tax.

Figure 3.5: Average tax pass-through



Notes: The estimated average pass-through at each month was calculated by dividing the coefficients ρ from the estimated regression (Equation 3.1) by the tax value (10%). Vertical spikes represent the 95% confidence intervals, with errors clustered at the brand level. The complete table with the raw coefficients from the regression estimated following Equation 3.1 can be found in Appendix C.1, Table C.1.

3.5.2 Last digit effect

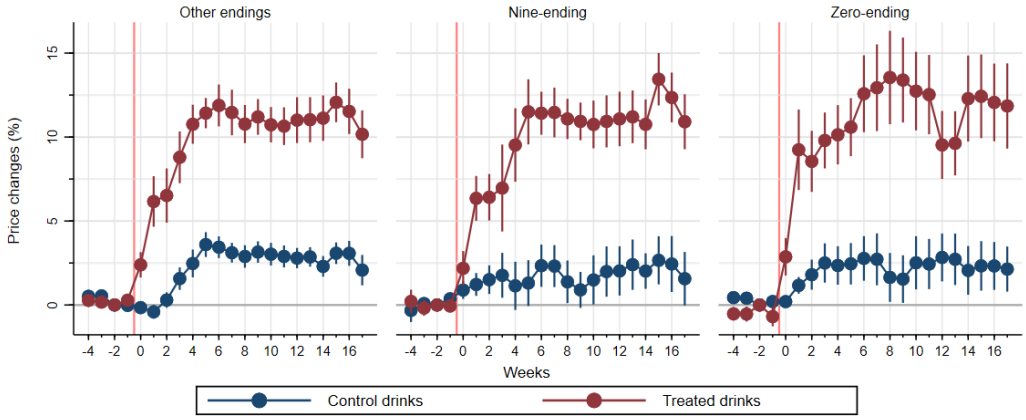
To analyse the role of price endings on tax pass-through, I divide items in 3 groups (j) based on the ending of their most frequent price in December 2020, what I call the *reference price*. Those are nine-ending, zero-ending, and other endings. I set *other endings* as the base category and evaluate whether reference price endings in nine or zero had any differential impact on price changes after the introduction of the tax. The estimation equation in this case, thus, includes an interaction of the variables of interest with the new variable denoting price endings:[7]

$$
\begin{aligned}
Log(P_{it}) &= \sum_{j=0}^{2} \beta_j \mathbb{1}\{Ending_i = j\} \\
&+ \sum_{j=0}^{2} \sum_{m=1}^{4} \gamma_{jm} \mathbb{1}\{Ending_i = j\} \times \mathbb{1}\{Month_t = m\} + \sum_{j=0}^{2} \lambda_j T_i \times \mathbb{1}\{Ending_i = j\} \\
&+ \sum_{j=0}^{2} \sum_{m=1}^{4} \rho_{jm} T_i \times \mathbb{1}\{Ending_i = j\} \times \mathbb{1}\{Month_t = m\} + \Omega' \mathrm{X}_{it} + \alpha_i + \epsilon_{it} \quad (3.2)
\end{aligned}
$$

Figure 3.6 shows the evolution of prices for each of the three ending groups among treated and control products. In all three cases, there are parallel trends the four weeks before the tax and a sharp increase in prices of taxed drinks in the first weeks of 2021. Figure 3.7, on the other hand, shows the estimated tax pass-through by reference price ending. Surprisingly, products initially priced with psychological endings (nine or zero) had a slightly larger pass-through than products initially priced with other endings, although the differences are not statistically significant at the 95% confidence level. Previous research in the literature of sticky prices had identified that products initially priced with psychological endings were less likely to update after a change in VAT (Knotek et al., 2019). My results could be driven by either the opposite phenomenon - that psychological prices were more likely to change after the VAT change in Spain -, or by a large price increase among the fewer products initially priced at psychological endings that did indeed update prices, even if the share of items actually changing prices was lower among treated products with nine and zero endings than among *other endings* ones.

[7]The specification in this page is written without a base category of price endings (j) for better readability and interpretation of the pass-through parameters ρ_{jm}.

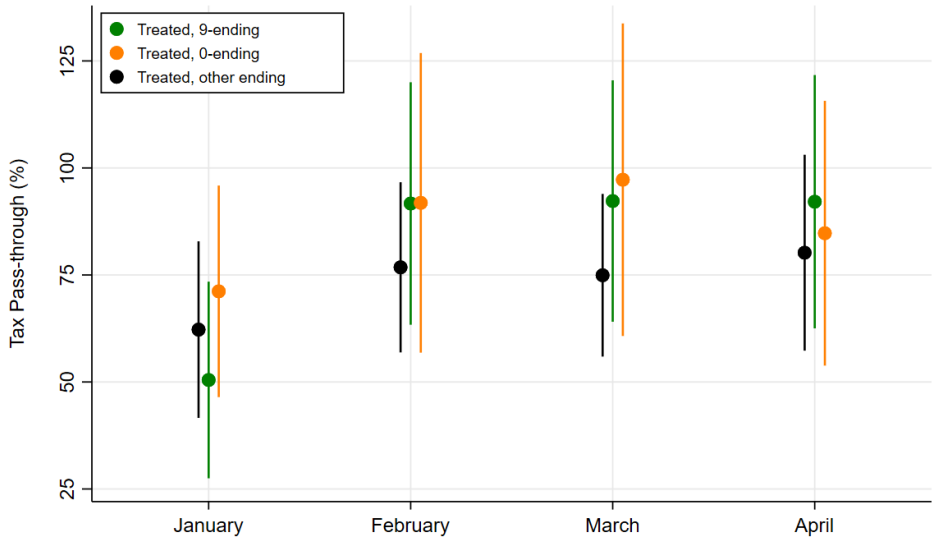
Figure 3.6: Evolution of prices before and after the tax change.



Notes: This graph plots the coefficients of a panel regression of the logarithm of prices on week dummies interacted with a dummy for each treatment group (taxed and untaxed) and reference price ending (nine, zero, or other), with item fixed effects. Vertical spikes represent 90% confidence intervals, with errors clustered at the brand level.

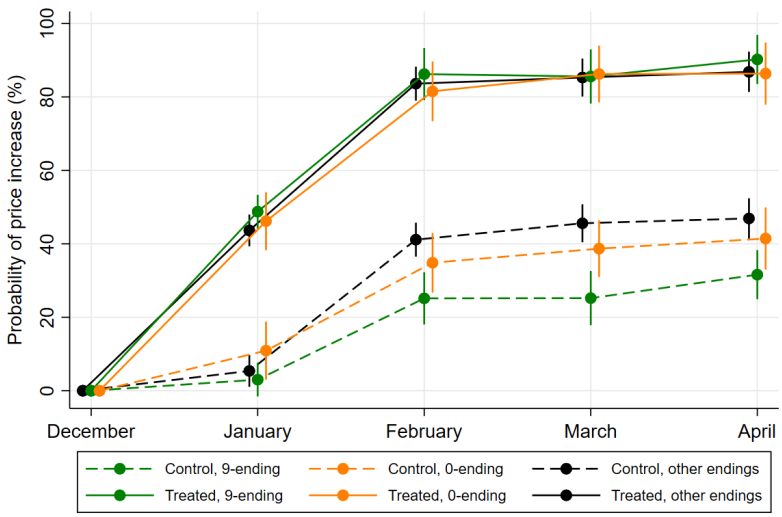
To further explore the dynamics behind the different levels of tax pass-through, I run an additional set of regressions that estimate the probability that a product increased its price from the initial reference price following the tax change. The explanatory variables are exactly the same as Equation 3.1, but in this case, the dependent variable is a dummy (D_{it}) that takes value one if the price of product i at time t is above its reference price (i.e. if it is above its December 2020 modal price). Figure 3.8 illustrates the estimated coefficients of a regression with a Panel Linear Probability Model (P-LPM). While one can observe that, after January 2021, products initially priced with nine and zero endings were less likely to increase in the control group, these differences do not show between treated drinks. Therefore, treated products were as likely to increase their prices regardless of their initial price ending, but those initially priced with zero endings increased by a larger amount than the ones initially priced with any other endings. Moreover, the larger pass-through estimate of products initially priced with nine-endings is somehow misleading since it is driven by the price stickiness of control drinks with the same ending. Indeed, taxed beverages initially priced with nine endings evolved very similarly to *other ending* treated products (Figure 3.6).

Figure 3.7: Average tax pass-through, by initial price ending



Notes: The estimated average pass-through at each month was calculated by dividing the coefficients ρ from the estimated regression (Equation 3.2) by the tax value (10%). Vertical spikes represent the 95% confidence intervals, with errors clustered at the brand level. The complete table with the raw coefficients from the regression estimated following Equation 3.2 can be found in Appendix C.1, Table C.2.

Figure 3.8: Price rigidity, by initial price ending



Notes: Coefficients from the panel regression estimated with a specification similar to Equation 3.1 but using the dummy D_{it} as the dependent variable, controlling for item fixed effects, were combined to produce this graph. The connected dots represent the increased probability of taxed and untaxed products to be priced above their most frequent (mode) price of December 2020, by month after the tax change. The vertical spikes represent 90% confidence intervals. Among control drinks, those with December's mode price ending in either nine or zero were less likely to increase in the following months. Among treated drinks, the probability of a product to be priced above its reference price from January onwards does not seem to differ by initial price ending. The complete table with raw values can be found in Appendix C.1, Table C.3.

3.5.3 First decimal effect

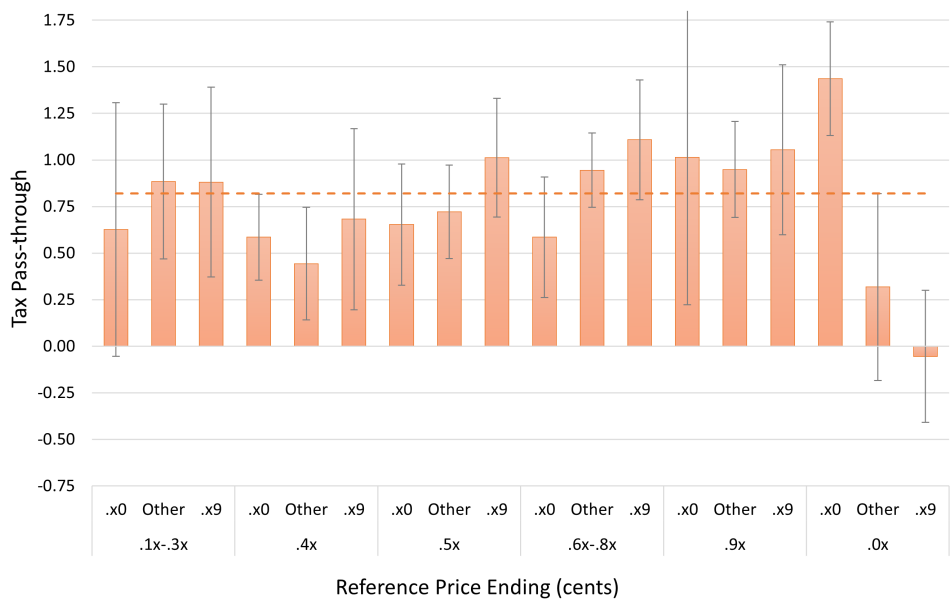
As mentioned in the previous section and illustrated in Figure 3.3, zero and nine endings are not necessarily independent of their preceding digit. For example, the most frequent two-digit ending among zero-ending prices is double zero, and the most frequent two-digit ending of all nine-ending prices is double nine. Both of these lie at each side of the round (integer) number, around which changes are expected to be most salient to consumers since they result in a change of the left digit.

In the previous set of regressions, I pooled all zero and nine endings together, irrespective of their preceding digit. This could lead to underestimating the effect of psychological endings if only double zero or double nine endings are relevant, but for example, endings in 20 or 29 are not. Therefore, I repeat the same exercise but introducing an additional set of interaction terms. To reduce the number of interactions, I classify all items based on whether the first decimal of their reference price (i.e. their modal price in December) fell within one of six options: zero (prices ending in .0x), one to three (prices ending in .1x, .2x, or .3x), four (.4x), five (.5x), six to eight (.6x, .7x, or .8x), and nine (.9x). The rationale behind this classification is to allow the highest flexibility to the digits around the two most relevant thresholds (full and half Euros) without adding an entire set of ten new dummies, one for each possible value.

Using k to denote each of the six new dummies capturing the first decimal digit of each product's reference price, the estimation equation for this section results in Equation 3.3. To simplify the notation, the interaction of price ending j (nine, zero, or other) with the first decimal category k is denoted by a single dummy for the combined ending kj. Furthermore, I move away from the specification with monthly dummies and, in this case, I simplify it to a dummy variable ($Post_t$) that denotes whether day t is after 31st December 2020. I exclude the transition period through which most prices were updated (January), and run this new set of regressions only using observations from December, February, March, and April. The dependent variable, Y_{it} is the logarithm of prices when estimating pass-through, and a dummy (D_{it}) with value one when the price of an item on day t is above its reference price when estimating price rigidity as the probability of price increases.

$$
\begin{aligned}
Y_{it} = & \sum_{k=1}^{6}\sum_{j=0}^{2} \beta_{kj} \mathbb{1}\{Ending_i = kj\} \\
& + \sum_{k=1}^{6}\sum_{j=0}^{2} \gamma_{kj} \mathbb{1}\{Ending_i = kj\} \times Post_t + \sum_{k=1}^{6}\sum_{j=0}^{2} \pi_{kj} \mathbb{1}\{Ending_i = kj\} \times T_i \\
& + \sum_{k=1}^{6}\sum_{j=0}^{2} \rho_{kj} \mathbb{1}\{Ending_i = kj\} \times T_i \times Post_t + \Omega' \mathrm{X}_{it} + \alpha_i + \epsilon_{it} \quad (3.3)
\end{aligned}
$$

Figure 3.9: Average tax pass-through, by initial price ending (two digits)

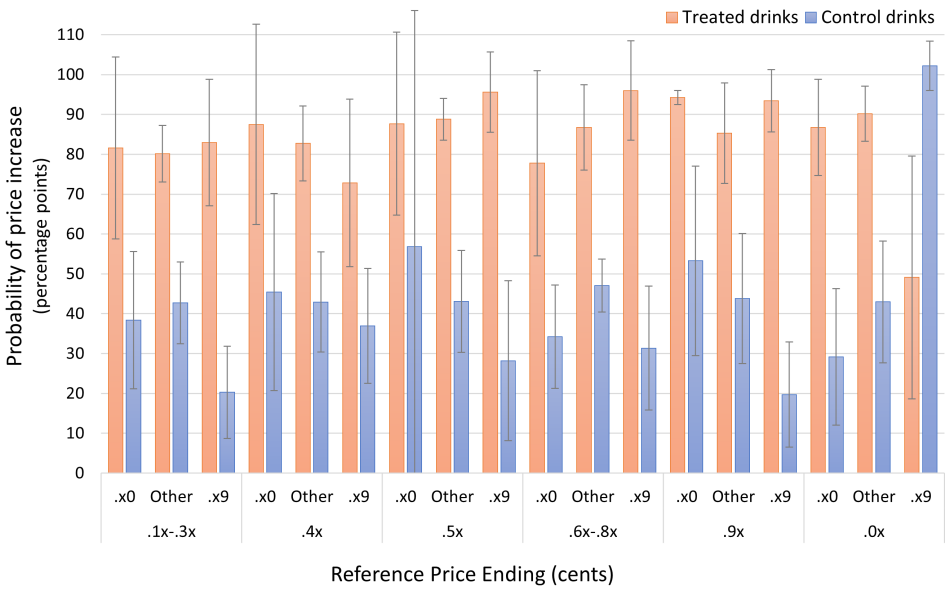


Notes: The estimated average pass-through at each month was calculated by dividing the coefficients ρ from the estimated regression (Equation 3.3) by the tax value (10%). The dependent variable, in this case, was $\log(P_{it})$. Vertical spikes represent the 90% confidence intervals, with errors clustered at the brand level. The complete table with the raw coefficients from the regression estimated following Equation 3.3 can be found in Appendix C.1, Table C.4.

The results of the panel regression used to estimate pass-through following Equation 3.3 are illustrated in Figure 3.9. These results can be better interpreted together with the estimated probabilities of products in 2021 to be priced above their reference price from December 2020 (Figure 3.10). Looking at zero-endings first, it seems that, while zero-endings reduce pass-through if preceded by any number different than nine or zero, double zero amplifies pass-through to a level that implies over-shifting of the tax to consumer prices. The effect of nine-endings seems to be less dependent on its preceding digit except for the ending in 09 cents (which represents a very small share of items). For all other nine-ending combinations, the introduction of the tax seems to have overcome the usual price rigidity of nine-endings that can be still observed among control products. On the other

hand, products with reference price endings within the range .40 to .58 seem to experience the lowest levels of pass-through, suggesting that halves of Euro may also set a relevant psychological threshold so far ignored by the literature.

Figure 3.10: Price rigidity, by initial price ending (two digits)



Notes: Coefficients from the panel regression estimated following the specification in Equation 3.3, using the dummy D_{it} as the dependent variable and controlling for item fixed effects, were combined to produce this graph. The connected dots represent the increased probability of taxed and untaxed products to be priced above their reference price (i.e. the most frequent (mode) price of December 2020) after the tax change. The vertical spikes represent 90% confidence intervals, with errors clustered at the brand level. The complete table with raw values can be found in Appendix C.1, Table C.5.

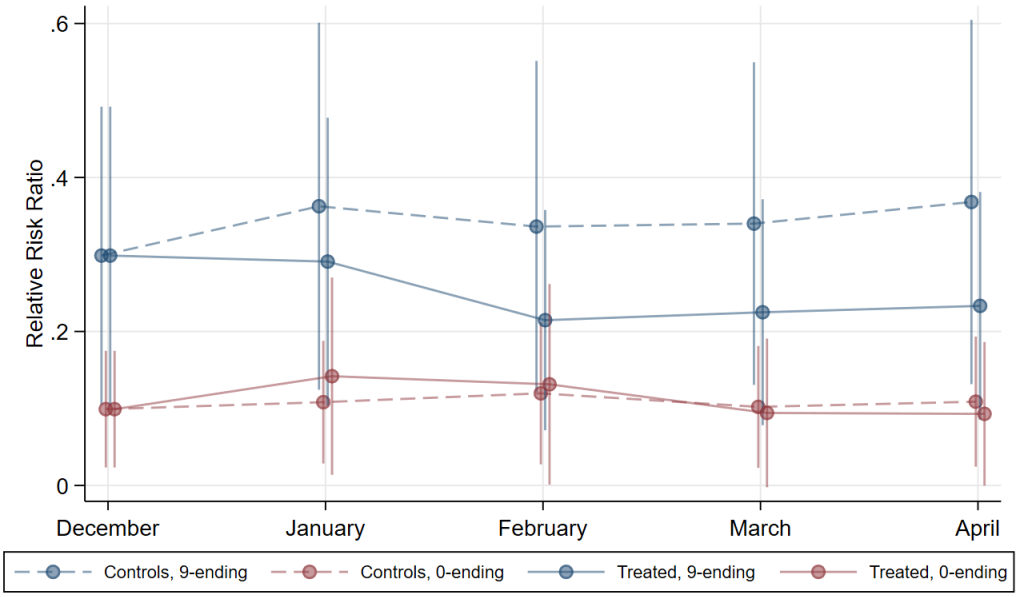
3.5.4 Strategic pricing

The last relevant question around psychological pricing and tax pass-through is whether sellers reacted to the new tax by using psychological prices more often in an attempt to reduce the reaction of consumers in response to price changes. In this case, I create a categorical variable capturing whether the price of item i at time t ends in nine, zero, or other endings, which I set as the dependent variable with three possible outcomes. Using a multinomial logistic regression, I estimate the conditional probability of an item to be priced at each of those endings by month and treatment group, controlling for the item's reference price (in logs) and seller-month fixed effects (both included in the control vector X_{it} in Equation 3.4). The results from the regression (Figure 3.11) do not show an increase in the relative probability of prices of treated items to end in either of the psychological endings after the tax change affecting treated products. Indeed, the reduction

in psychological endings was larger among treated products than among control drinks. Therefore, it does not seem that sellers resorted to psychological endings to attenuate the effect of the tax on consumption.

$$Ending_{it} = \sum_{m=0}^{4} \beta_m \mathbb{1}\{Month_t = m\} + \sum_{m=0}^{4} \gamma_m T_i \times \mathbb{1}\{Month_t = m\} + \Gamma' \mathrm{X}_{it} + \nu_{it} \quad (3.4)$$

Figure 3.11: Evolution of psychological endings (last digit)



Notes: Coefficients from a multinomial logistic regression based on the estimation equation 3.4 were combined to produce this graph. Each point represents the estimated relative probability of a treated or a control item to be priced at each of the two evaluated psychological endings, compared to the probability of being priced at non-psychological endings (*other*). The vertical spikes represent 90% confidence intervals. The full table with raw values can be found in Appendix C.1, Table C.6.

3.6 Conclusion

The Spanish market of soft drinks exhibited a relevant degree of psychological pricing by the end of 2020, before the increase of the value-added tax (VAT) on sweetened drinks. Despite the role of psychological endings on price stickiness identified by previous studies, I estimate a nearly full pass-through of the tax on sweetened sodas using waters, beers, and vegetable milks as controls, within two months after the tax change. Further analysis of the evolution of prices does not indicate that sellers recurred to psychological endings to pass a larger tax burden on to consumers while reducing consumption responses. Nevertheless, psychological endings of initial prices did indeed influence tax pass-through, with products initially priced ending in nine or zero exhibiting the largest increases in prices after the tax change. The difference was especially remarkable for products initially priced at round (double zero) endings, where the tax was over-shifted on consumer prices. Another relevant finding is that halves of Euro also seem to play a relevant role as psychological thresholds, with products initially priced at endings between .40 and .58 exhibiting lower pass-through. This is of utmost relevance to the literature of psychological pricing, which so far has only paid attention to nine and zero endings (Levy et al., 2011; Ater and Gerlitz, 2017; Knotek et al., 2019).

Furthermore, my results highlight that the role of psychological thresholds when shocks to the supply curve (marginal cost) are as large as 10 per cent of the final price may differ from their role under smaller shocks. While nine-ending prices seem to be more rigid among control drinks, which did not suffer any increase in taxes, treated products with initial prices ending in nine were as likely or more to increase after the introduction of the tax than those with other initial price endings. Moreover, the last specification of my analysis, interacting the two last digits of prices (cents), highlights that the effect of the last price digit is significantly affected by its preceding digit.

Finally, the findings of this paper could contribute to explain the wide variety of estimates of sugar tax pass-through in the literature if they were replicated using data from other countries. The frequency of round endings, for example, could generate large differences across jurisdictions or types of retailers, and policy-makers should take it into account when predicting who will bear the burden of the tax. Moreover, my findings show that, although psychological endings have been previously found to increase price stickiness of grocery prices, soda taxes that account for ten per cent of the selling price seem to overcome such rigidity, supporting the recommendation of the WHO that suggests that sugar

taxes ought to be large to be effective in triggering consumption responses.

Unfortunately, the lack of consumption data does not allow to test whether consumers effectively respond more to changes of products initially priced with a psychological ending than those priced at other endings. Moreover, in countries with soda taxes per litre or ounce, items with the same price would bear different taxes based on their volume. This would allow for a more precise estimation of how the effect of psychological endings changes depending on the size of the shock to marginal costs. Unfortunately, this is not possible in the Spanish setting with a proportional tax independent of item sizes. Both analyses are the natural next steps of this research agenda.

Appendix A (for Chapter 1)

A.1 Experiment design

This appendix provides further details of the experiment and its reward structure, as well as some sample screenshots.

Table A.1: Randomisation checks

	Treatment 1	**Treatment 2**	**Control**
Males	52%	52%	49%
Age, years	36.4	36.74	37.36
	(12.14)	(11.76)	(12.25)
Higher Education	72%	72%	75%
Employed	77%	77%	80%
Ever filed taxes	89%	90%	92%
Earnings in real life	$ 67,512	$ 51,627	$ 52,575
	(372021)	(58204)	(54033)
Taxes in real life	$ 6,703	$ 6,288	$ 6,633
	(14205)	(12450)	(10159)
Quintile in real life	3.6	3.55	3.55
	(1.06)	(1.09)	(1.09)
Focus measure	0.89	0.90	0.89
	(0.16)	(0.16)	(0.16)
Correct tasks (/12)	4.68	4.86	4.78
	(2.39)	(2.52)	(2.36)
Matched to High Ref. Group	52%	51%	49%

Notes: This table provides summary statistics of socio-demographic and performance variables to check whether the random allocation of participants to treatment groups yielded balanced samples. Only the subsample used in the analysis was used for the calculations: participants with a focus measure above 25% and solving correctly between one and ten real effort tasks. Higher education refers to university level (Bachelor or above). For continuous variables, the standard deviation is reported in parenthesis.

A.1.1 Earning scheme and incentives

Participants earned money from three different sources:

1. **Participation:** all participants that completed the experiment were awarded a participation fee of 0.40 USD.

2. **Net earned income**: amount earned in the initial task ($10,000$ ECU for each correct answer), minus taxes. The equivalent to real currency is $50,000$ ECU to 1 USD.

Participants in the two treatment groups were told the tax scheme they were facing when their earned income was revealed, while participants on the control group were told their earnings would be subject to a tax, but the system in place was only revealed at the end of the experiment.

3. **Bonus income**: additional tax-free amount (25,000 ECU) earned for correctly answering the questions regarding income distribution statistics. Since all three questions are related, the answer to one question could condition another one. To prevent this, a lottery decides which one of the three questions is chosen for the bonus payment.

The random matching to 99 other players was achieved through a random match of the player to one of 2 pre-existing groups of 99 previous records. Matching to 99 pre-loaded previous records is an alternative to a 100 player simultaneous game, which would introduce additional issues (waiting time, potential drop-outs, etc.). Each of those groups was formed by bootstrapping from a sample of pre-existing records from the trial phase. At a point in the experiment (see Figure 1.8), participants were shown a selected subsample of 9 of those players in their matched group as a source of partial information to update their beliefs.

A.1.2 Choice of tax schemes

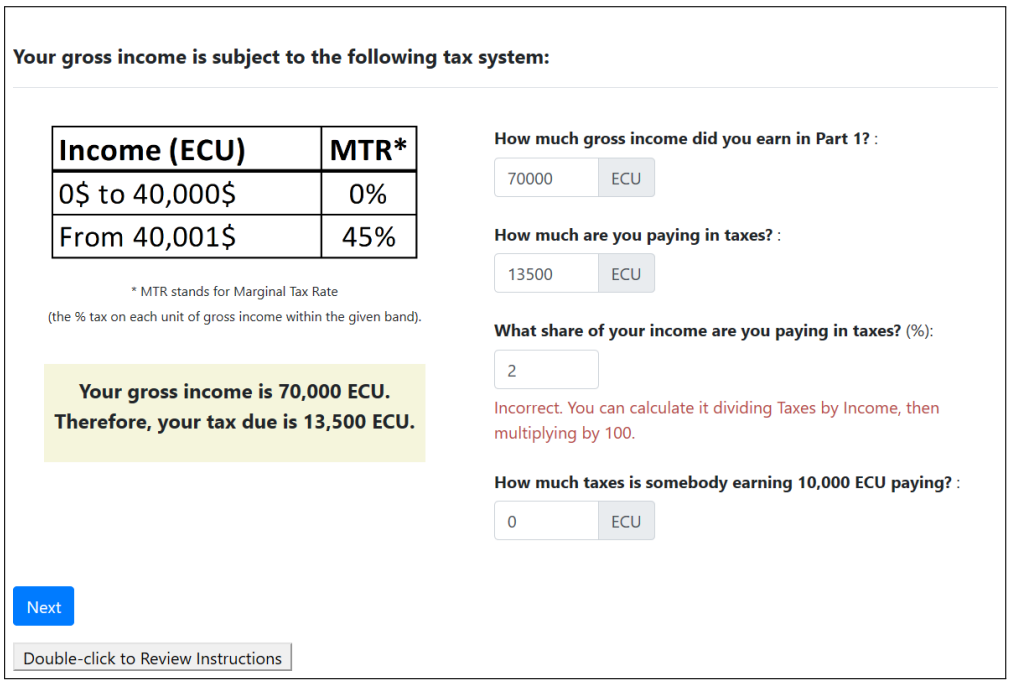
The number of brackets, income thresholds and marginal tax rates for the progressive tax schedule were chosen taking the UK Income Tax as a main reference.[1] To ensure that participants looked carefully at the tax schedule, they were asked a few simple details of the tax system in place.

A.1.3 Real effort tasks

The task was partially inspired by Sutter and Weck-Hannemann (2003) and is similar in spirit to Gneezy et al. (2003), remaining simpler than Chow (1983). It consists of solving four mathematical operations and eight scrambled words (four names of countries and four names of animals), and it is preceded by a trial task involving one question of each type so that players become acquainted with the dynamics. Each word/operation has a time limit of five seconds to be solved, and every correct answer yields 10,000ECU.

[1]Details are published on-line by the HMRC: *https://www.gov.uk/government/publications/rates-and-allowances-income-tax/income-tax-rates-and-allowances-current-and-past*.

Figure A.1: Income Tax Information for Treatment 2 (progressive system)



Your gross income is subject to the following tax system:

Income (ECU)	MTR*
0$ to 40,000$	0%
From 40,001$	45%

* MTR stands for Marginal Tax Rate
(the % tax on each unit of gross income within the given band).

Your gross income is 70,000 ECU. Therefore, your tax due is 13,500 ECU.

How much gross income did you earn in Part 1? :
70000 ECU

How much are you paying in taxes? :
13500 ECU

What share of your income are you paying in taxes? (%):
2
Incorrect. You can calculate it dividing Taxes by Income, then multiplying by 100.

How much taxes is somebody earning 10,000 ECU paying? :
0 ECU

Next

Double-click to Review Instructions

For the purpose of the experiment, the earning task needs to follow three main requirements: incomplete information on the income distribution, realism, and low dispersion. The need for incomplete information and realism is obvious, while low dispersion helps to minimize income effects (making respondents more comparable); it also helps with planning the budget of the experiment.

Incomplete information was ensured since all players knew the reward structure, probability of multiplier, time limitations and the lower and upper income bounds, but they did not know other players' performance. Although an upper bound does usually not exist in reality, it is unavoidable in an experiment.

A.1.4 Instructions

Since this experiment was not distributed in a lab, the instructions had to be fully displayed on the screen, and participants were not provided with a printed copy. The set of initial instructions was available for review on every screen through a button at the bottom of the page.

Screenshots of the instructions at each step are provided below, with self-explanatory headings:

Figure A.2: Examples of real effort tasks

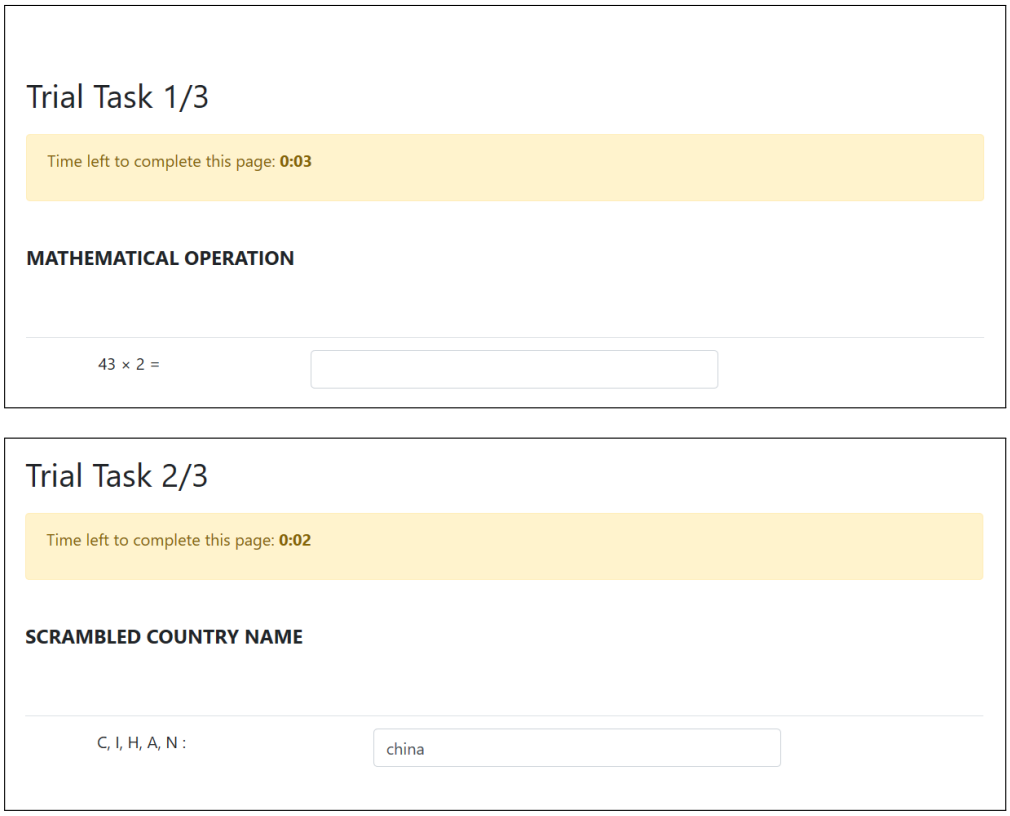


Figure A.3: Initial page of the Experiment

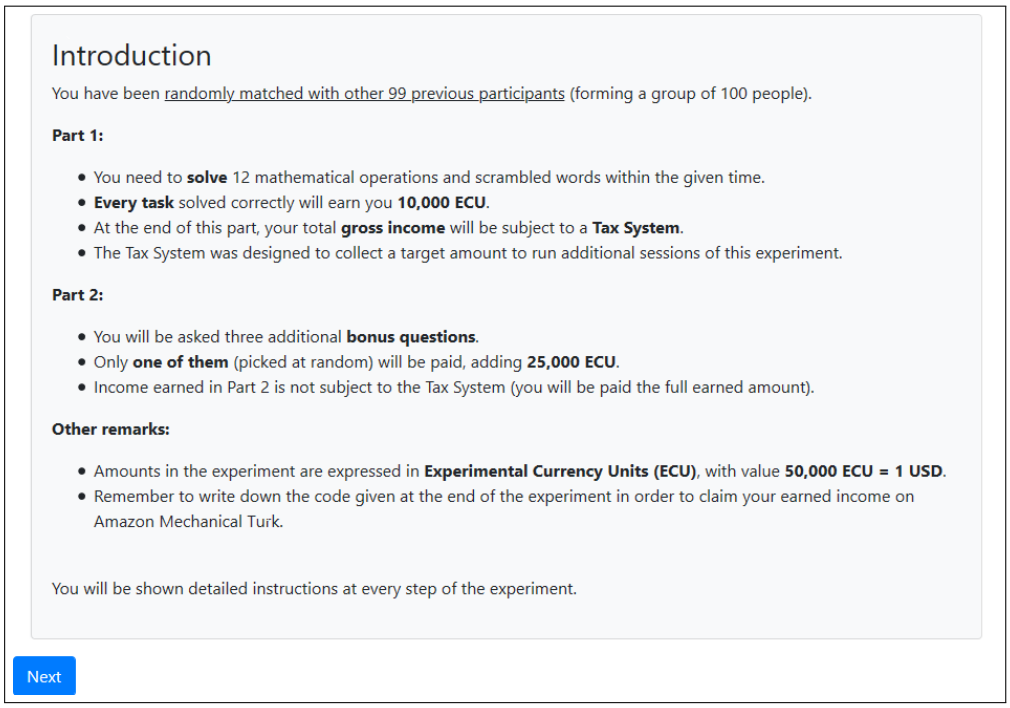


Figure A.4: Instructions for Part 1 (real effort tasks)

Part 1

You are required to solve some **short mathematical operations and scrambled names** of countries and animals.

- For every task, you will be told if it is a mathematical operation, scrambled animal name or scrambled country name.
- You will have **5 seconds** to answer each task by typing the number or ordered letters in the textbox.
- Your answer will be **automatically submitted after timeout** (5 seconds) and a new task will be shown.
- You can **submit your answer before timeout** by pressing the *Enter* key on your keyboard.
- **Every correct answer has a value of 10,000 ECU.**

You will start with a trial of 3 tasks and at the end you will be shown your trial results.

The **trial section does not award any income** and is for practice only.

After the trial, you will begin your actual 12 tasks.

You may want to use paper and pen, or a calculator.

Next

Double-click to Review Instructions

Figure A.5: Instructions for Part 2 (elicitation of beliefs)

Part 2 - Bonus Questions

Remember you were randomly matched to 99 other players at the beginning of the experiment, forming a group of 100.

You will now be asked **3 questions** related to the earnings of that group of 100 people in Part 1.

A random draw will define which question will be paid at the end of the experiment.

If **Questions 1 or 2** are selected by the random draw, you will be paid the bonus if your answer is correct.

If **Question 3** is the one selected, the probability you select as an answer will define whether your payment depends on the statement in Question 3 being true, or on a random lottery*.

This payment mechanism is designed so that answering honestly every question gives you higher probability of earning the bonus.

The bonus amount is **25,000 ECU**.

*Learn more about the payment method for Question 3.

Next

Figure A.6: Additional information about the payment structure

*Learn more about the payment method for Question 3.

How is your payment determined if Question 3 is randomly selected for payment?

- The computer will generate a lottery box with 100 balls, with a random proportion of purple balls (from 1 to 100).
  Note: all proportions of purple balls are equally likely (have the same probability of happening).
- Your **reported probability** defines if your bonus will depend on the statement in Question 3 being indeed true, or on a random draw from the lottery box. How?
  - If the box has a higher number of purple balls than your reported probability:
    A ball will be randomly drawn. If it is **purple, you win**.
  - If the box has a lower or equal number of purple balls than your reported probability:
    The statement in Question 3 is checked. If it was indeed **true, you win**.
- Your bonus depends on the lottery only when the probability of wining it is higher than your reported probability of Question 3 being true. Therefore, the method applied will always favour your probability to earn the bonus.

Example:

- Imagine you **reported a probability of 20 percent** that the statement in Question 3 is true.
- The computer randomly generated a box with 100 balls, 40 of which are purple.
- This means the probability to randomly draw a purple ball from the box is **40 percent**.
  Since 40 (purple balls) is higher than 20 (reported probability):
  - **Your payment depends on the box**, and it does no longer matter whether the statement in Q3 is true or false.
  - A ball will be randomly drawn from that box: **if it is purple, you win** the bonus. Otherwise, you don't.

- Imagine instead that you had **reported a probability of 60 percent** that the statement in Question 3 is true.
- This time, the computer randomly generated a box with 100 balls, 50 of which are purple.
  Since 60 (reported probability) is higher than 50 (purple balls):
  - **Your payment depends on the statement** in Question 3, and the lottery box no longer matters in this case.
  - The statement in Question 3 is checked: **if it is true, you win** the bonus. Otherwise, you don't.

Next

Appendix B (for Chapter 2)

B.1 Additional context tables

Table B.1: List of countries with soda taxes

Country	From	To	Products affected	Min. sugar content	Tax value
American Samoa	2001		Excise for locally produced and import tax		
Bahrain	2017		All soft-drinks and energy drinks (all aerated beverages)	No minimum	50% on soft drinks 100% on energy drinks
Barbados	2015		All sugary drinks, except pure natural juice, milk based, waters and powdered drinks	No minimum Any sugar or flavouring	10% excise tax pre VAT
Belgium	2016		All non-alcoholic sweetened beverages	No minimum Any sweetener or flavouring	0.068 €/l. (initially 0.03 €/l.) Intermediary liquids: 0.41 €/l.
Bermuda	2018		All non-alcoholic sweetened beverages, except fruit and vegetable juices	No minimum Any sweetener or flavouring	50% VAT on drinks 50% VAT on syrups
Brunei	2017		All sugar-sweetened beverages, except milk-based drinks and fruit juices	Sweetened & coffee > 6g / 100ml Soya milk > 7g / 100ml Malted or chocolate > 8g / 100ml	0.40 B$/l.
Chile	2014		All non-alcoholic sugar-sweetened beverages, except water, milk-based and pure juices	Any sugar or flavouring Additional if > 6.25g / 100ml	Initial common excise tax: 13% Increased for sugary: 18% Decreased for non/low sug: 10%
Cook Islands	2014		All non-alcoholic beverages	> 5 g. / 100ml	VAT on drinks
Denmark	2011	2012	Saturated fat products	> 2.3g saturated fat / 100g fat	50% VAT
Dominica	2015		All food and drinks with high sugar content	No minimum	$ 0.20 /l. 10% on energy drinks
Ecuador	2016		Any sweetened soft drinks and all energy drinks, except dairy products, minueral water, and >50% fruit juices	No minimum Additional if > 2.5g / 100ml	Basic 10% excise tax Special rate $ 0.25 /100g sugar (initially $ 0.18 /100g sugar)
Estonia	2018		All non-alcoholic beverages with sweeteners or sugar, except 100% juice drinks and milk drinks.	> 5g. / 100ml.	Only artificial sweet.: 0.10€/l. Only sugar, 5-8g.: 0.10€/l. Artif. + 5-8g sugar: 0.20€/l. > 8g. Sugar: 0.30€/l.
Ethiopia	2020		All sweetened beverages, except fruit and vegetable juices	No minimum Any sweetener	25%
Fiji	2011		All sweetened beverages and flavoured or coloured sugar syrups, except milk-based drinks	No minimum Any sweetener or flavouring	Imported: 15% excise tax (initially 10%) Local product: 35cts/l. (initially 10cts/l.)
Finland	2011	2017	All non-alcoholic beverage containing sweeteners or flavours, except small producers	No minimum Additional if >0.5% sugar	Basic 0.11 €/l. Additional + 0.21 €/l. (in 2014: + 0.11 €/l.)
France	2012		All sweetened non-alcoholic beverages	Proportional	Initially 0.075 €/l. Up to 0.20 €/l.
French Polynesia	2002		All sweetened drinks and beer	No minimum Any sweetener	Imported: 60 cfp/l. Local product: 40 cfp/l.
Hungary	2011		All non-alcoholic sweetened beverages	Sweetened drinks > 8g/100ml No minimum for energy drinks	Soft drinks: 15 huf/l. (initially 5 huf/l.) Energy drinks: 300 huf/l. (Initially 250 huf/l.)
India	2017		All sweetened or flavoured beverages	No minimum	Additional +12% on GST
Ireland	2018		All non-alcoholic sugar-sweetened drinks, except fruit juices and dairy products	> 5g / 100ml Higher if > 8g / 100ml	6g-8g : 0.20 €/l. > 8g : 0.30 €/l.
Kiribati	2014		All non-alcoholic sweetened or flavoured beverages, excep fruit and vegetable juices	No minimum	40% excise tax
Latvia	2004		All non-alcoholic sweetened or flavoured beverages, except fruit and vegetable juices and nectars	No minimum	0.74 €/l. (initially 0.285 €/l.)
Malaysia	2019		All sugary beverages	Sweetened drinks > 5g/100ml Natural drinks > 12g/100ml	RM 0.40 /l.
Mauritius	2013		All sugar-sweetened beverages	No minimum	MUR 0.06 /g. of sugar (initially MUR 0.03 /g.)
Mexico	2014		All sugary beverages, except milks and yoghurts	No minimum Additional if > 275cal / 100g.	Basic: MXN 1 /l. Energy drinks: 25% excise tax Caloric drinks: 8% excise tax
Morocco	2020		All sugary non-alcoholic beverages	No minimum Additional if < 10% pure juice	Basic: MAD 1-1.5 /l. VAT Add. < 5g./100ml : MAD 3 /l. Add. 5g.-10g. : MAD 3.75 /l. Add. > 10g. : MAD 4.5 /l.
Norway	1981		All sweetened non-alcoholic beverages	No minimum Additional for sugar-sweetened	Artificial sweet.: 2.54 NOK /l. > 8g./100ml : 3.63 NOK /l. (In 2017: 3.34 NOK /l.) (In 2018: 4.75 NOK /l.)

(Continues in the next page)

List of countries with soda taxes (continued)

Country	From	To	Products affected	Min. sugar content	Tax value
Oman	2019		All sweetened beverages in any form (ready to drink, concentrate, powder, gel), except pure vegetable and fruit juices, dairy products (>75% milk), and dietary sup.	No minimum	50% excise tax
Peru	2018		All sugary non-alcoholic beverages, including 0% alcohol beer	No minimum Additional for > 6g / 100ml	Basic: 17% excise tax Additional: 25% excise tax
Philippines	2018		All sweetened beverages (unless stevia or coco sugar), except milk, pure juices, coffee and tea	No minimum	Regular: 6 pesos /l. If using corn syrup: 12 pesos /l.
Portugal	2017		All sweetened non-alcoholic beverages (SSB&ASB), except milk and alternatives, and fruit juices	No minimum Higher for > 8g / 100ml	< 8g / 100ml. : 0.08 €/l. > 8g / 100ml. : 0.16 €/l.
Qatar	2019		All soft drinks	No minimum	Carbonated drinks: 50% tax Energy drinks: 100% tax
Samoa	1984		All soft drinks	No minimum	0.4 talas /l. (initially 0.3 talas /l.)
Saudi Arabia	2017		All carbonated drinks and energy drinks	No minimum	Carbonated drinks: 50% tax Energy drinks: 100% tax
Seychelles	2019		All sugary drinks, except fresh locally produced drinks and plain milk	> 5g / 100ml	4 SR / l.
South Africa	2018		All sweetened beverages except pure natural juices	> 4g / 100ml	0.021 SGD / gram above 4 g.
Spain (Catalonia)	2017		All sugary and (caloric) sweetened soft drinks, except fruit juices	> 5g / 100ml	5g.-8g. / 100ml. : 0.08 €/l. > 8g. : 0.12 €/l.
Spain (all regions)	2021		All sweetened beverages unless served to consume at restaurants and bars. Milk-based products are exempt	No minimum	Additional +11% VAT
Sri Lanka	2017		All sugar-sweetened beverages	No minimum	RS 0.30 / g. of sugar (initially RS 0.50 / g.)
St Helena	2014		All carbonated drinks containing sugar	> 1.5g / 100ml	0.75 SHP /l.
Thailand	2017		All soft drinks containing sugar.	> 6g / 100ml	Sweetened drinks: 14% VAT Fruit&Veg. juices: 10% VAT 6g. - 8g. / 100ml. : 0.10 baht /l. 8 - 10g. / 100ml. : 0.30 baht /l. 10 - 14g. / 100ml. : 0.50 baht /l. > 14g. / 100ml. : 1 baht / l. (doubling every 2 years)
Tonga	2013		All soft drinks containing sugar or sweeteners	No minimum	1.5 TOP / l. (initially 0.5 TOP /l.)
UAE	2017		All carbonated drinks and energy drinks, except unflavoured aerated water	No minimum	Carbonated drinks: 50% Energy drinks: 100%
UK	2018		Pre-packaged soft drinks with added sugar, except pure fruit juices and milk substitutes	> 5g / 100ml	5g. - 8g. / 100ml. : £ 0.18 /l. > 8g. / 100ml. : £ 0.24 /l.
USA (S. Francisco)	2018		All sugar-sweetened beverages except except pure natural juices and milk substitutes	> 7cal / 100ml	$ 0.01 /oz.
USA (Berkeley)	2015		All (caloric) sweetened beverages	No minimum	$ 0.01 /oz.
USA (Navajo N.)	2015		All minimal-to-non-nutritional food items	No minimum	2%
USA (Albany)	2017		All (caloric) sweetened beverages, except milk products and pure natural juices	No minimum	$ 0.01 /oz.
USA (Philadelphia)	2017		All non-alcoholic sweetened beverages, except milk products and pure natural juices	No minimum	$ 0.015 /oz.
USA (Boulder)	2017		All (caloric) sweetened beverages, except milk products and alcoholic beverages	> 1.4g / 100ml	$ 0.02 /oz.
USA (Oakland)	2017		All sugar-sweetened beverages, except milk products and pure natural juices	> 7cal / 100ml	$ 0.01 /oz.
USA (Seattle)	2018		All (caloric) sweetened beverages, except milk products and pure natural juices	No minimum	$ 0.0175 / oz.
USA (Washington)	2010	2010	All carbonated drinks	No minimum	$ 0.0175 / oz. (initially $ 0.0016 / oz.)
Vanuatu	2015		All sweetened carbonated beverages	No minimum	50 vatu /l.

Source: Elaborated by the author.

Table B.2: Changes in levels of sugar 2017-2018, by product

Group	Matrix Brand	Product	Sugar content (g/100ml) Previous (2017)	New formula (2018)
AG Barr	Irn-Bru	Irn-Bru	10.3	4.7
AG Barr	Rockstar	Rockstar Energy Drink	15.6	4.8
AG Barr	Rubicon	Rubicon	12	4.8
Britvic	7UP	7UP	10.6	7
Britvic	Britvic	Britvic Ginger Ale	9.1	9.1
Britvic	Britvic	Britvic tonic	5.8	4.7
Britvic	J2O	J2O Fruit Blends	7.3	4.7
Britvic	Lipton	Lipton Ice Tea	6.5	4.5
Britvic	Mountain Dew	Mountain Dew	13	12
Britvic	Pepsi	Pepsi original	11	11
Britvic	Robinsons	Robinsons Barley water	4.4	4.4
Britvic	Robinsons	Robinsons Fruit Creations	0	0
Britvic	Robinsons	Robinsons Squash'd	0	0
Britvic	Tango	Tango	4.3	4.3
Coca-Cola	Coca-Cola	Coca-cola original	10.6	10.6
Coca-Cola	Coca-Cola	Coke cherry	11.2	11.2
Coca-Cola	Dr Pepper	Dr Pepper	7.2	4.9
Coca-Cola	Fanta	Fanta Lemon	8.3	4.5
Coca-Cola	Fanta	Fanta Orange	6.9	4.6
Coca-Cola	Fanta	Fanta Other	6.3	4.8
Coca-Cola	Lilt	Lilt	4.6	4.6
Coca-Cola	Schweppes	Schweppes Indian Tonic	5.1	4.9
Coca-Cola	Schweppes	Schweppes Lemonade	4.2	4.2
Coca-Cola	Sprite	Sprite	6.6	3.3
Fever-Tree	Fever-Tree	Fever-Tree Indian Tonic	8	2.9
Lucozade	Lucozade Energy	Lucozade original	8.7	4.5
Monster	Monster	Monster Energy drink	11	11
Monster	Monster	Monster Juiced	8.4	8.4
Old Jamaica	Old Jamaica	Old Jamaica Ginger beer	15.2	4.9
Red Bull	Red Bull	Red Bull	11	11
Ribena	Ribena	Ribena Squash Concentrate	10	4.6
San Pellegrino	San Pellegrino	San Pellegrino Lemon	7	4.6
Vimto	Vimto	Vimto Original Squash	9.1	4.6

Source: Elaborated by the author.
Notes: Sugar levels that would fall within the lower and the higher levies are highlighted in yellow (lighter) and orange (darker), respectively.

B.2 Additional regression tables

Table B.3: Tax pass-through of the SDIL and spillover on substitute drinks (full)

VARIABLES	(1) Quarterly	(2) Quarterly
Pass-through at Q1	1.149***	1.037***
	(0.204)	(0.289)
Pass-through at Q2	1.100***	1.090***
	(0.185)	(0.208)
Pass-through at Q3	0.886**	0.943***
	(0.323)	(0.308)
Pass-through at Q4	1.389***	1.397***
	(0.195)	(0.175)
Spillover on Sugar-Free (Q1)	0.316***	0.329**
	(0.068)	(0.149)
Spillover on Sugar-Free (Q2)	0.536***	0.521***
	(0.098)	(0.133)
Spillover on Sugar-Free (Q3)	0.295**	0.334**
	(0.111)	(0.137)
Spillover on Sugar-Free (Q4)	0.508***	0.512***
	(0.119)	(0.139)
Spillover on Other Carbonated (Q1)	0.270**	-0.051
	(0.104)	(0.225)
Spillover on Other Carbonated (Q2)	0.209	0.140
	(0.192)	(0.179)
Spillover on Other Carbonated (Q3)	-0.259	-0.063
	(0.154)	(0.141)
Spillover on Other Carbonated (Q4)	0.167	0.198
	(0.198)	(0.180)
Spillover on Non-Carbonated (Q1)	0.039	0.241
	(0.175)	(0.199)
Spillover on Non-Carbonated (Q2)	-0.115	-0.329
	(0.195)	(0.290)
Spillover on Non-Carbonated (Q3)	-0.225	-0.150
	(0.237)	(0.252)
Spillover on Non-Carbonated (Q4)	0.140	0.094
	(0.246)	(0.230)
Observations	23,831	78,273
R-squared	0.249	0.091
Number of ASIN	41	134
Item FE	YES	YES
Season-Category FE	NO	YES

Robust standard errors in parenthesis. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.
Notes: These two columns contain the full set of coefficients for the regression estimated following Equation 2.11. Column (2) corresponds to Column (3) of Table 2.6, including the initially omitted coefficients for *Other Carbonated* and *Non-Carbonated* drinks.

Appendix C (for Chapter 3)

C.1 Regression tables

Table C.1: Regression table for Figure 3.5 - Average tax pass-through

VARIABLES	(1) OLS	(2) OLS	(3) Panel
January (β_1)	-0.019**	-0.006	0.000
	(0.009)	(0.007)	(0.005)
February ($beta_2$)	0.014	0.050***	0.040***
	(0.011)	(0.011)	(0.008)
March (β_3)	0.038**	0.058***	0.036***
	(0.016)	(0.018)	(0.008)
April (β_4)	0.069***	0.072***	0.040***
	(0.011)	(0.018)	(0.008)
Treated (λ_1)	-0.020	0.066	
	(0.151)	(0.152)	
Treated # January (ρ_1)	0.085***	0.071***	0.062***
	(0.012)	(0.011)	(0.009)
Treated # February (ρ_2)	0.094***	0.067***	0.082***
	(0.013)	(0.015)	(0.009)
Treated # March (ρ_3)	0.070***	0.052**	0.082***
	(0.022)	(0.025)	(0.009)
Treated # April (ρ_4)	0.051**	0.036*	0.083***
	(0.020)	(0.021)	(0.010)
Observations	119,546	119,546	119,546
R-squared	0.002	0.052	0.2117
Seller-month FE	NO	YES	YES
Product FE	NO	NO	YES
Number of id			1,075

Robust standard errors in parenthesis. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.
Notes: This table shows the estimates from a set of regressions following the specification in Equation 3.1. The coefficients ρ_{jm} in column (3), calculated through a panel regression with item fixed effects, are the estimates of tax pass-through depicted in Figure 3.5. Errors were clustered at the brand level.

Table C.2: Regression table for Figure 3.7 - Last digit effect (price changes)

VARIABLES	(1) LPM	(2) LPM	(3) Panel
Treated (λ_0)	0.074	0.144	
	(0.160)	(0.158)	
Base Category			
January (γ_{01})	-0.020**	-0.006	-0.002
	(0.010)	(0.008)	(0.005)
February (γ_{02})	0.017**	0.049***	0.043***
	(0.009)	(0.009)	(0.007)
March (γ_{03})	0.051***	0.064***	0.039***
	(0.013)	(0.014)	(0.007)
April (γ_{04})	0.078***	0.076***	0.041***
	(0.013)	(0.017)	(0.008)
Treated × January (ρ_{01})	0.084***	0.069***	0.062***
	(0.013)	(0.012)	(0.011)
Treated × February (ρ_{02})	0.091***	0.066***	0.077***
	(0.011)	(0.011)	(0.010)
Treated × March (ρ_{03})	0.057***	0.045**	0.075***
	(0.017)	(0.020)	(0.010)
Treated × April (ρ_{04})	0.056***	0.045**	0.080***
	(0.021)	(0.022)	(0.012)
Psychological Endings			
Nine-ending ($\beta_1 - \beta_0$)	0.080	0.095	
	(0.110)	(0.105)	
Zero-ending ($\beta_2 - \beta_0$)	0.624***	0.482***	
	(0.148)	(0.138)	
Treated × Nine-ending ($\lambda_1 - \lambda_0$)	0.006	0.001	
	(0.147)	(0.138)	
Treated × Zero-ending ($\lambda_2 - \lambda_0$)	-0.606***	-0.522***	
	(0.203)	(0.190)	
Nine-ending × January ($\gamma_{11} - \gamma_{01}$)	-0.005	0.001	0.007
	(0.017)	(0.016)	(0.006)
Nine-ending × February ($\gamma_{12} - \gamma_{02}$)	0.022	0.018	-0.014
	(0.026)	(0.028)	(0.010)
Nine-ending × March ($\gamma_{13} - \gamma_{03}$)	0.006	0.002	-0.015
	(0.033)	(0.034)	(0.009)
Nine-ending × April ($\gamma_{14} - \gamma_{04}$)	-0.032	-0.014	-0.005
	(0.038)	(0.038)	(0.010)
Zero-ending × January ($\gamma_{21} - \gamma_{01}$)	0.019	0.010	0.013***
	(0.019)	(0.021)	(0.005)
Zero-ending × February ($\gamma_{22} - \gamma_{02}$)	-0.031	-0.014	0.001
	(0.021)	(0.018)	(0.009)
Zero-ending × March ($\gamma_{23} - \gamma_{03}$)	-0.066**	-0.031	0.001
	(0.030)	(0.025)	(0.009)
Zero-ending × April ($\gamma_{24} - \gamma_{04}$)	-0.049**	-0.037*	0.005
	(0.023)	(0.021)	(0.008)
Treated × Nine-ending × January ($\rho_{11} - \rho_{01}$)	-0.018	-0.025	-0.012
	(0.021)	(0.021)	(0.013)
Treated × Nine-ending × February ($\rho_{12} - \rho_{02}$)	-0.048	-0.050	0.015
	(0.034)	(0.036)	(0.016)
Treated × Nine-ending × March ($\rho_{13} - \rho_{03}$)	-0.060	-0.068	0.017
	(0.048)	(0.050)	(0.014)
Treated × Nine-ending × April ($\rho_{14} - \rho_{04}$)	-0.039	-0.054	0.012
	(0.050)	(0.051)	(0.015)
Treated × Zero-ending × January ($\rho_{21} - \rho_{01}$)	0.010	0.023	0.009
	(0.025)	(0.027)	(0.015)
Treated × Zero-ending × February ($\rho_{22} - \rho_{02}$)	0.060***	0.052**	0.015
	(0.023)	(0.020)	(0.019)
Treated × Zero-ending × March ($\rho_{23} - \rho_{03}$)	0.125***	0.102***	0.022
	(0.036)	(0.032)	(0.019)
Treated × Zero-ending × April ($\rho_{24} - \rho_{04}$)	0.043	0.035	0.005
	(0.030)	(0.029)	(0.018)
Observations	119,546	119,546	119,546
R-squared	0.030	0.068	0.214
Seller-month FE	NO	YES	YES
Product FE	NO	NO	YES
Number of id			1,075

Robust standard errors in parenthesis. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.
Notes: This table shows the estimates from a set of regressions following the specification in Equation 3.2 but with *other endings* as the base category. Therefore, the coefficients in this table for the other two endings categories denote the differential value from the base category. Errors were clustered at the brand level. For an easier interpretation, I have added the correspondence with the notation in Equation 3.2. The points in Figure 3.7 correspond to the ρ parameters as described in Equation 3.2 divided by 10% (to be expressed as price changes in proportion to the tax change) and were calculated through a linear combination of the values in column (3) of this table.

Table C.3: Regression table for Figure 3.8 - Last digit effect (price rigidity)

VARIABLES	(1) LPM	(2) LPM	(3) Panel
Treated (λ_0)	0.006	0.004	
	(0.013)	(0.012)	
Base Category			
January (γ_{01})	0.123***	0.055*	0.057**
	(0.018)	(0.028)	(0.028)
February (γ_{02})	0.393***	0.416***	0.414***
	(0.037)	(0.041)	(0.041)
March (γ_{03})	0.439***	0.460***	0.460***
	(0.038)	(0.040)	(0.039)
April (γ_{04})	0.419***	0.462***	0.473***
	(0.039)	(0.040)	(0.040)
Treated × January (ρ_{01})	0.353***	0.387***	0.382***
	(0.048)	(0.045)	(0.046)
Treated × February (ρ_{02})	0.432***	0.428***	0.429***
	(0.051)	(0.051)	(0.053)
Treated × March (ρ_{03})	0.408***	0.404***	0.404***
	(0.053)	(0.052)	(0.055)
Treated × April (ρ_{04})	0.436***	0.425***	0.407***
	(0.056)	(0.056)	(0.057)
Psychological Endings			
Nine-ending ($\beta_1 - \beta_0$)	-0.007	-0.008	
	(0.009)	(0.008)	
Zero-ending ($\beta_2 - \beta_0$)	0.002	0.007	
	(0.010)	(0.010)	
Treated × Nine-ending ($\lambda_1 - \lambda_0$)	-0.005	-0.005	
	(0.011)	(0.011)	
Treated × Zero-ending ($\lambda_2 - \lambda_0$)	-0.024	-0.027*	
	(0.015)	(0.015)	
Nine-ending × January ($\gamma_{11} - \gamma_{01}$)	-0.016	-0.032	-0.024
	(0.025)	(0.027)	(0.029)
Nine-ending × February ($\gamma_{12} - \gamma_{02}$)	-0.159***	-0.165***	-0.160***
	(0.050)	(0.050)	(0.048)
Nine-ending × March ($\gamma_{13} - \gamma_{03}$)	-0.201***	-0.209***	-0.203***
	(0.048)	(0.048)	(0.045)
Nine-ending × April ($\gamma_{14} - \gamma_{04}$)	-0.129**	-0.145***	-0.151***
	(0.051)	(0.051)	(0.048)
Zero-ending × January ($\gamma_{21} - \gamma_{01}$)	0.030	0.050	0.050
	(0.039)	(0.033)	(0.033)
Zero-ending × February ($\gamma_{22} - \gamma_{02}$)	-0.113*	-0.078	-0.069
	(0.058)	(0.054)	(0.055)
Zero-ending × March ($\gamma_{23} - \gamma_{03}$)	-0.119*	-0.082	-0.076
	(0.063)	(0.058)	(0.058)
Zero-ending × April ($\gamma_{24} - \gamma_{04}$)	-0.122*	-0.071	-0.061
	(0.067)	(0.058)	(0.057)
Treated × Nine-ending × January ($\rho_{11} - \rho_{01}$)	0.082	0.087*	0.076*
	(0.053)	(0.046)	(0.045)
Treated × Nine-ending × February ($\rho_{12} - \rho_{02}$)	0.190***	0.192***	0.186***
	(0.063)	(0.063)	(0.063)
Treated × Nine-ending × March ($\rho_{13} - \rho_{03}$)	0.203***	0.207***	0.202***
	(0.063)	(0.063)	(0.061)
Treated × Nine-ending × April ($\rho_{14} - \rho_{04}$)	0.160***	0.168***	0.181***
	(0.060)	(0.060)	(0.059)
Treated × Zero-ending × January ($\rho_{21} - \rho_{01}$)	-0.050	-0.078	-0.068
	(0.062)	(0.056)	(0.058)
Treated × Zero-ending × February ($\rho_{22} - \rho_{02}$)	0.006	-0.018	-0.018
	(0.074)	(0.072)	(0.074)
Treated × Zero-ending × March ($\rho_{23} - \rho_{03}$)	0.026	0.000	0.011
	(0.074)	(0.071)	(0.071)
Treated × Zero-ending × April ($\rho_{24} - \rho_{04}$)	0.011	-0.022	-0.016
	(0.074)	(0.070)	(0.069)
Observations	119,546	119,546	119,546
R-squared	0.384	0.398	0.398
Seller-month FE	NO	YES	YES
Product FE	NO	NO	YES
Number of id			1,075

Robust standard errors in parenthesis. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.
Notes: This table shows the estimates from a set of Linear Probability Model regressions following a specification similar to Equation 3.2 but with the dummy D_{it} as the dependent variable, and with *other endings* as the base category. Therefore, the coefficients in this table for the other two endings categories denote the differential value from the base category. Errors were clustered at the brand level. For an easier interpretation, I have added the correspondence with the notation in Equation 3.2. The points in Figure 3.8 were calculated through a linear combination of the values in column (3) of this table. December is set to zero deducting the items' fixed effects.

Table C.4: Regression table for Figure 3.9 - First decimal effect (price changes)

VARIABLES	(1) OLS	(2) OLS	(3) Panel
After Tax Change (γ_{40})	0.030**	0.045***	0.037***
	(0.014)	(0.013)	(0.007)
Ending .1x-.3x × Post ($\gamma_{10} - \gamma_{40}$)	0.014	0.017	0.001
	(0.016)	(0.015)	(0.006)
Ending .4x × Post ($\gamma_{20} - \gamma_{40}$)	0.061**	0.063**	0.012
	(0.029)	(0.027)	(0.012)
Ending .5x × Post ($\gamma_{30} - \gamma_{40}$)	0.009	0.005	0.002
	(0.016)	(0.015)	(0.010)
Ending .9x × Post ($\gamma_{50} - \gamma_{40}$)	0.027	0.032	-0.001
	(0.031)	(0.029)	(0.012)
Ending .0x × Post ($\gamma_{60} - \gamma_{40}$)	0.006	0.009	0.041
	(0.039)	(0.037)	(0.032)
Ending .x9 × Post ($\gamma_{41} - \gamma_{40}$)	-0.008	0.003	-0.014
	(0.024)	(0.027)	(0.014)
Ending .x0 × Post ($\gamma_{42} - \gamma_{40}$)	-0.056**	-0.038**	-0.011
	(0.023)	(0.018)	(0.013)
Nine-ending in .1x-.3x × Post ($\gamma_{11} - \gamma_{10}$)	-0.084*	-0.094**	-0.007
	(0.047)	(0.047)	(0.019)
Zero-ending in .1x-.3x × Post ($\gamma_{12} - \gamma_{10}$)	0.042	0.048	0.043
	(0.053)	(0.049)	(0.034)
Nine-ending in .4x × Post ($\gamma_{21} - \gamma_{20}$)	-0.011	-0.025	0.042
	(0.041)	(0.043)	(0.029)
Zero-ending in .4x × Post ($\gamma_{22} - \gamma_{20}$)	-0.022	-0.012	0.009
	(0.059)	(0.054)	(0.020)
Nine-ending in .5x × Post ($\gamma_{31} - \gamma_{30}$)	0.096	0.081	0.002
	(0.095)	(0.086)	(0.023)
Zero-ending in .5x × Post ($\gamma_{32} - \gamma_{30}$)	-0.001	0.008	0.018
	(0.037)	(0.041)	(0.017)
Nine-ending in .9x × Post ($\gamma_{51} - \gamma_{50}$)	0.013	0.009	-0.005
	(0.072)	(0.073)	(0.027)
Zero-ending in .9x × Post ($\gamma_{52} - \gamma_{50}$)	0.026	0.019	0.022
	(0.048)	(0.046)	(0.024)
Nine-ending in .0x × Post ($\gamma_{61} - \gamma_{60}$)	0.017	0.012	0.002
	(0.041)	(0.042)	(0.031)
Zero-ending in .0x × Post ($\gamma_{62} - \gamma_{60}$)	0.002	0.001	-0.030
	(0.055)	(0.048)	(0.031)
After Tax Change (Post) × Treated (ρ_{40})	0.101***	0.094***	0.094***
	(0.021)	(0.021)	(0.012)
Ending .1x-.3x × Post × Treated ($\rho_{10} - \rho_{40}$)	-0.010	-0.024	-0.006
	(0.047)	(0.047)	(0.026)
Ending .4x × Post × Treated ($\rho_{20} - \rho_{40}$)	-0.125***	-0.137***	-0.050***
	(0.041)	(0.043)	(0.019)
Ending .5x × Post × Treated ($\rho_{30} - \rho_{40}$)	-0.020	-0.024	-0.022
	(0.043)	(0.044)	(0.016)
Ending .9x × Post × Treated ($\rho_{50} - \rho_{40}$)	-0.041	-0.052	0.000
	(0.048)	(0.049)	(0.017)
Ending .0x × Post × Treated ($\rho_{60} - \rho_{40}$)	-0.045	-0.055	-0.063*
	(0.043)	(0.040)	(0.034)
Ending .x9 × Post × Treated ($\rho_{41} - \rho_{40}$)	-0.016	-0.022	0.016
	(0.034)	(0.039)	(0.022)
Ending .x0 × Post × Treated ($\rho_{42} - \rho_{40}$)	0.034	0.017	-0.036*
	(0.056)	(0.055)	(0.020)
Nine-ending in .1x-.3x × Post × Treated ($\rho_{11} - \rho_{10}$)	0.017	0.016	-0.017
	(0.092)	(0.094)	(0.046)
Zero-ending in .1x-.3x × Post × Treated ($\rho_{12} - \rho_{10}$)	0.008	0.015	0.010
	(0.092)	(0.092)	(0.055)
Nine-ending in .4x × Post × Treated ($\rho_{21} - \rho_{20}$)	-0.075	-0.081	0.007
	(0.109)	(0.118)	(0.038)
Zero-ending in .4x × Post × Treated ($\rho_{22} - \rho_{20}$)	0.069	0.078	0.050
	(0.076)	(0.072)	(0.033)
Nine-ending in .5x × Post × Treated ($\rho_{31} - \rho_{30}$)	-0.103	-0.094	0.013
	(0.104)	(0.097)	(0.032)
Zero-ending in .5x × Post × Treated ($\rho_{32} - \rho_{30}$)	-0.035	-0.026	0.029
	(0.065)	(0.073)	(0.031)
Nine-ending in .9x × Post × Treated ($\rho_{51} - \rho_{50}$)	0.017	0.014	-0.006
	(0.081)	(0.084)	(0.035)
Zero-ending in .9x × Post × Treated ($\rho_{52} - \rho_{50}$)	0.026	0.031	0.043
	(0.090)	(0.090)	(0.049)
Nine-ending in .0x × Post × Treated ($\rho_{61} - \rho_{60}$)	-0.030	-0.041	-0.054
	(0.052)	(0.053)	(0.044)
Zero-ending in .0x × Post × Treated ($\rho_{62} - \rho_{60}$)	0.116	0.118*	0.148***
	(0.071)	(0.066)	(0.042)
Observations	93,507	93,507	93,507
R-squared	0.101	0.134	0.000332
Seller-month FE	NO	YES	NO
Product FE	NO	NO	YES
Number of id			1,075

Robust standard errors in parenthesis. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.
Notes: The regressions in this table follow the specification in Equation 3.2 but with *Endings in .1x-.3x* ($j = 0$) and *Other Endings* ($k = 4$) as the base categories. Errors were clustered at the brand level. The correspondence of each coefficient with the notation in Equation 3.2 was added for easier interpretation. The level of pass-through can be calculated by dividing the parameters ρ_{kj} by 10% (the tax increase). Figure 3.9 uses the coefficients from column (3).

Table C.5: Regression table for Figure 3.10 - First decimal effect (price rigidity)

VARIABLES	(1) LPM	(2) LPM	(3) Panel
After Tax Change (γ_{40})	0.452***	0.470***	0.470***
	(0.040)	(0.040)	(0.040)
Ending .1x-.3x × Post ($\gamma_{10} - \gamma_{40}$)	-0.058	-0.047	-0.043
	(0.055)	(0.055)	(0.057)
Ending .4x × Post ($\gamma_{20} - \gamma_{40}$)	-0.041	-0.037	-0.041
	(0.079)	(0.079)	(0.075)
Ending .5x × Post ($\gamma_{30} - \gamma_{40}$)	-0.045	-0.050	-0.039
	(0.076)	(0.078)	(0.077)
Ending .9x × Post ($\gamma_{50} - \gamma_{40}$)	-0.046	-0.037	-0.032
	(0.105)	(0.107)	(0.099)
Ending .0x × Post ($\gamma_{60} - \gamma_{40}$)	-0.045	-0.042	-0.041
	(0.093)	(0.090)	(0.090)
Ending .x9 × Post ($\gamma_{41} - \gamma_{40}$)	-0.154	-0.162	-0.157
	(0.100)	(0.099)	(0.096)
Ending .x0 × Post ($\gamma_{42} - \gamma_{40}$)	-0.186**	-0.137*	-0.128
	(0.092)	(0.081)	(0.080)
Nine-ending in .1x-.3x × Post ($\gamma_{11} - \gamma_{10}$)	-0.062	-0.069	-0.068
	(0.136)	(0.134)	(0.131)
Zero-ending in .1x-.3x × Post ($\gamma_{12} - \gamma_{10}$)	0.116	0.084	0.085
	(0.142)	(0.130)	(0.126)
Nine-ending in .4x × Post ($\gamma_{21} - \gamma_{20}$)	0.113	0.102	0.097
	(0.158)	(0.157)	(0.155)
Zero-ending in .4x × Post ($\gamma_{22} - \gamma_{20}$)	0.117	0.156	0.153
	(0.161)	(0.154)	(0.152)
Nine-ending in .5x × Post ($\gamma_{31} - \gamma_{30}$)	0.000	-0.000	0.008
	(0.143)	(0.142)	(0.142)
Zero-ending in .5x × Post ($\gamma_{32} - \gamma_{30}$)	0.238	0.263	0.265
	(0.335)	(0.383)	(0.380)
Nine-ending in .9x × Post ($\gamma_{51} - \gamma_{50}$)	-0.060	-0.066	-0.084
	(0.148)	(0.148)	(0.143)
Zero-ending in .9x × Post ($\gamma_{52} - \gamma_{50}$)	0.271	0.240	0.222
	(0.184)	(0.187)	(0.184)
Nine-ending in .0x × Post ($\gamma_{61} - \gamma_{60}$)	0.747***	0.741***	0.749***
	(0.138)	(0.136)	(0.132)
Zero-ending in .0x × Post ($\gamma_{62} - \gamma_{60}$)	-0.012	-0.012	-0.010
	(0.147)	(0.141)	(0.132)
After Tax Change (Post) × Treated (ρ_{40})	0.385***	0.394***	0.397***
	(0.076)	(0.074)	(0.077)
Ending .1x-.3x × Post × Treated ($\rho_{10} - \rho_{40}$)	0.011	-0.016	-0.023
	(0.089)	(0.088)	(0.090)
Ending .4x × Post × Treated ($\rho_{20} - \rho_{40}$)	0.015	-0.006	0.001
	(0.102)	(0.104)	(0.105)
Ending .5x × Post × Treated ($\rho_{30} - \rho_{40}$)	0.083	0.072	0.060
	(0.099)	(0.102)	(0.104)
Ending .9x × Post × Treated ($\rho_{50} - \rho_{40}$)	0.043	0.035	0.018
	(0.121)	(0.120)	(0.114)
Ending .0x × Post × Treated ($\rho_{60} - \rho_{40}$)	0.100	0.084	0.076
	(0.118)	(0.118)	(0.120)
Ending .x9 × Post × Treated ($\rho_{41} - \rho_{40}$)	0.236*	0.254*	0.249*
	(0.132)	(0.137)	(0.139)
Ending .x0 × Post × Treated ($\rho_{42} - \rho_{40}$)	0.095	0.042	0.038
	(0.149)	(0.146)	(0.145)
Nine-ending in .1x-.3x × Post × Treated ($\rho_{11} - \rho_{10}$)	0.004	0.002	0.003
	(0.187)	(0.190)	(0.191)
Zero-ending in .1x-.3x × Post × Treated ($\rho_{12} - \rho_{10}$)	-0.036	0.029	0.019
	(0.258)	(0.240)	(0.239)
Nine-ending in .4x × Post × Treated ($\rho_{21} - \rho_{20}$)	-0.282	-0.287	-0.289
	(0.224)	(0.236)	(0.238)
Zero-ending in .4x × Post × Treated ($\rho_{22} - \rho_{20}$)	-0.030	-0.009	-0.016
	(0.206)	(0.203)	(0.202)
Nine-ending in .5x × Post × Treated ($\rho_{31} - \rho_{30}$)	-0.014	-0.023	-0.032
	(0.183)	(0.189)	(0.194)
Zero-ending in .5x × Post × Treated ($\rho_{32} - \rho_{30}$)	-0.210	-0.180	-0.187
	(0.347)	(0.409)	(0.408)
Nine-ending in .9x × Post × Treated ($\rho_{51} - \rho_{50}$)	0.080	0.046	0.073
	(0.172)	(0.174)	(0.170)
Zero-ending in .9x × Post × Treated ($\rho_{52} - \rho_{50}$)	-0.071	-0.065	-0.043
	(0.244)	(0.246)	(0.245)
Nine-ending in .0x × Post × Treated ($\rho_{61} - \rho_{60}$)	-1.190***	-1.205***	-1.253***
	(0.286)	(0.290)	(0.272)
Zero-ending in .0x × Post × Treated ($\rho_{62} - \rho_{60}$)	0.070	0.078	0.065
	(0.208)	(0.207)	(0.205)
Observations	93,507	93,507	93,507
R-squared	0.450	0.462	0.459
Seller-month FE	NO	YES	NO
Product FE	NO	NO	YES
Number of id			1,075

Robust standard errors in parenthesis. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.
Notes: The regressions in this table follow the specification in Equation 3.2, but with the dummy D_{it} as dependent variable and *Endings in .1x-.3x* ($j = 0$) and *Other Endings* ($k = 4$) as base categories. Errors were clustered at the brand level. The correspondence with the notation in Equation 3.2 was added for easier interpretation. Figure 3.10 uses the coefficients from column (3).

Table C.6: Regression table for Figure 3.11 - Strategic Pricing

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	9-Ending	0-Ending	9-Ending	0-Ending	9-Ending	0-Ending
January ($\beta_1 - \beta_0$)	0.972	0.968	1.220**	1.107	1.219**	1.109
	(0.045)	(0.057)	(0.096)	(0.134)	(0.096)	(0.135)
February ($\beta_2 - \beta_0$)	0.978	0.988	1.129	1.222	1.128	1.223
	(0.079)	(0.102)	(0.154)	(0.215)	(0.154)	(0.218)
March ($\beta_3 - \beta_0$)	0.993	0.980	1.143	1.056	1.143	1.056
	(0.096)	(0.114)	(0.171)	(0.224)	(0.170)	(0.226)
April ($\beta_4 - \beta_0$)	0.918	1.103	1.243	1.118	1.242	1.116
	(0.095)	(0.116)	(0.208)	(0.237)	(0.208)	(0.238)
Treated (γ_0)	1.090	1.167	1.132	1.333	1.138	1.341
	(0.208)	(0.305)	(0.214)	(0.346)	(0.212)	(0.347)
Treated times January ($\gamma_1 - \gamma_0$)	0.910	1.354	0.810	1.305	0.811	1.301
	(0.143)	(0.332)	(0.124)	(0.341)	(0.124)	(0.341)
Treated times February ($\gamma_2 - \gamma_0$)	0.698*	1.182	0.651**	1.095	0.652**	1.093
	(0.147)	(0.368)	(0.138)	(0.359)	(0.138)	(0.360)
Treated times March ($\gamma_3 - \gamma_0$)	0.717	0.921	0.667**	0.898	0.667**	0.898
	(0.147)	(0.321)	(0.134)	(0.329)	(0.134)	(0.329)
Treated times April ($\gamma_4 - \gamma_0$)	0.731	0.845	0.635**	0.834	0.635**	0.836
	(0.141)	(0.275)	(0.130)	(0.287)	(0.131)	(0.286)
Constant (β_0)	0.284***	0.243***	0.272***	0.161***	0.308**	0.103***
	(0.039)	(0.044)	(0.048)	(0.028)	(0.156)	(0.062)
Observations	119,546		119,546		119,546	
Seller-month FE	NO		YES		YES	
Controls	NO		NO		YES	
Pseudo R-squared	0.00334		0.0240		0.0247	

Robust standard errors in parenthesis. Significance levels: *** p< 0.01, ** p< 0.05, * p< 0.1.
Notes: This table shows the coefficients of a multinomial logistic regression following Equation 3.4 but with December as the base month, as odds ratios (denoting the relative probability compared to *other endings*). The correspondence of the coefficients with the notation used in Equation 3.4 is added for easier interpretation. Errors were clustered at the brand level. Figure 3.11 uses the coefficients from columns (7-8).

C.2 Additional distribution tables

Table C.7: Distribution of the first decimal of prices in December 2020

		First decimal									
Category	**Seller**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0**
Cola Sodas	**Alcampo**	3%	9%	3%	18%	21%	12%	2%	6%	17%	9%
	Eroski	1%	2%	3%	4%	29%	27%	2%	0%	5%	27%
	Mercadona	0%	9%	9%	0%	32%	25%	0%	9%	17%	0%
Lemon Sodas	**Alcampo**	2%	2%	14%	16%	14%	11%	5%	7%	21%	9%
	Eroski	14%	10%	10%	26%	10%	10%	0%	0%	5%	17%
	Mercadona	10%	0%	0%	20%	10%	30%	10%	0%	0%	20%
Orange Sodas	**Alcampo**	4%	11%	13%	21%	11%	6%	7%	7%	7%	14%
	Eroski	5%	0%	23%	27%	0%	18%	0%	0%	9%	18%
	Mercadona	0%	0%	25%	0%	0%	50%	0%	0%	0%	25%
Energy Drinks	**Alcampo**	8%	6%	3%	3%	2%	0%	5%	13%	33%	27%
	Eroski	0%	9%	5%	14%	0%	0%	0%	0%	23%	50%
	Mercadona	13%	0%	13%	13%	0%	0%	25%	13%	25%	0%
Sport Drinks	**Alcampo**	0%	0%	9%	21%	15%	14%	10%	10%	7%	15%
	Eroski	13%	0%	0%	3%	28%	3%	20%	7%	3%	23%
	Mercadona	0%	0%	0%	29%	14%	0%	29%	0%	0%	29%
Tonic Water	**Alcampo**	4%	13%	17%	28%	12%	10%	3%	9%	4%	0%
	Eroski	12%	6%	0%	23%	36%	12%	6%	6%	0%	0%
	Mercadona	17%	0%	0%	21%	17%	0%	29%	0%	17%	0%
Sparkling Water	**Alcampo**	13%	9%	9%	0%	3%	13%	14%	8%	19%	13%
	Eroski	16%	9%	13%	6%	0%	10%	0%	0%	42%	4%
	Mercadona	33%	33%	0%	0%	0%	0%	0%	16%	17%	0%
Still Water	**Alcampo**	14%	9%	11%	20%	17%	11%	5%	4%	5%	2%
	Eroski	2%	8%	5%	24%	21%	18%	12%	6%	2%	2%
	Mercadona	8%	15%	8%	31%	0%	8%	0%	8%	15%	8%
Vegetable Milk	**Alcampo**										
	Eroski	5%	5%	8%	15%	12%	7%	11%	2%	24%	11%
	Mercadona	11%	0%	6%	16%	0%	11%	6%	28%	6%	16%
Alcoholic Beer	**Alcampo**	4%	6%	6%	12%	17%	17%	15%	10%	7%	4%
	Eroski	8%	8%	14%	9%	11%	14%	9%	10%	11%	6%
	Mercadona	2%	10%	16%	8%	12%	10%	10%	14%	9%	10%
Overall	**Alcampo**	6%	7%	9%	17%	15%	11%	7%	8%	12%	8%
	Eroski	6%	6%	9%	13%	15%	14%	8%	6%	11%	12%
	Mercadona	7%	7%	9%	14%	9%	12%	9%	12%	10%	10%

Notes: Each cell represents the proportion of prices within a seller-category pair with a given digit in the first decimal position. Different shades of yellow were used to highlight cells with a concentration of prices above the uniform distribution value (10%). In the bottom three rows, green colour was used to highlight the digits with a frequency above 10% in each supermarket pooling all categories together.

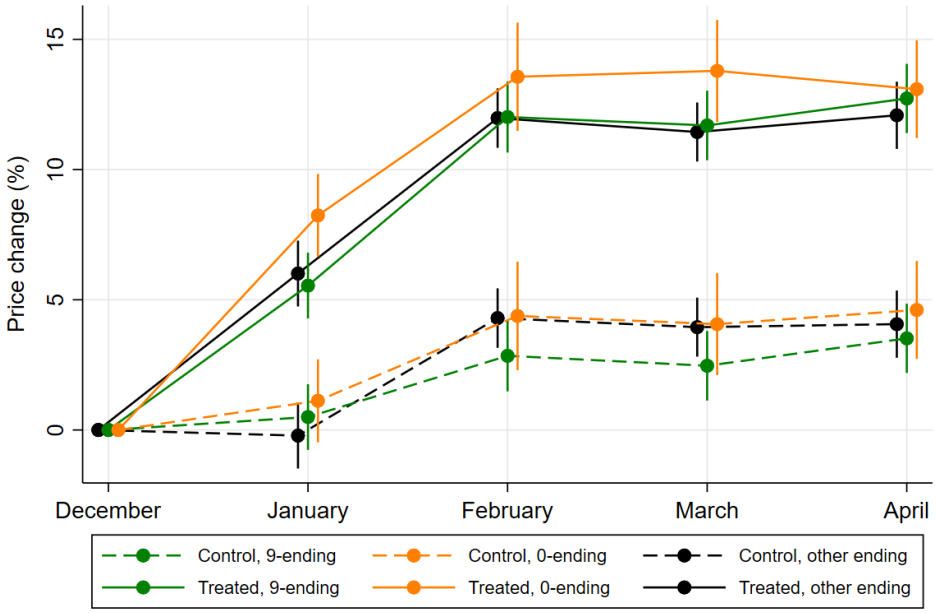
Table C.8: Distribution of the second decimal of prices in December 2020

		Second Decimal									
Category	**Seller**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0**
Cola Sodas	**Alcampo**	1%	13%	2%	9%	17%	11%	3%	11%	21%	13%
	Eroski	0%	6%	8%	19%	25%	9%	0%	4%	14%	16%
	Mercadona	0%	0%	0%	0%	17%	17%	0%	8%	6%	51%
Lemon Sodas	**Alcampo**	3%	13%	9%	9%	8%	2%	14%	7%	23%	13%
	Eroski	0%	14%	0%	14%	26%	0%	0%	0%	29%	17%
	Mercadona	0%	20%	0%	10%	10%	0%	0%	0%	10%	50%
Orange Sodas	**Alcampo**	4%	4%	4%	14%	18%	7%	3%	11%	21%	16%
	Eroski	11%	11%	0%	5%	25%	0%	0%	5%	32%	11%
	Mercadona	0%	71%	0%	0%	4%	0%	0%	0%	0%	25%
Energy Drinks	**Alcampo**	6%	5%	10%	14%	10%	5%	0%	10%	17%	22%
	Eroski	14%	9%	5%	0%	5%	0%	0%	0%	32%	36%
	Mercadona	0%	25%	0%	13%	25%	0%	13%	0%	13%	13%
Sport Drinks	**Alcampo**	7%	15%	0%	20%	1%	5%	1%	7%	24%	20%
	Eroski	0%	0%	7%	0%	54%	0%	0%	0%	13%	26%
	Mercadona	0%	0%	0%	0%	43%	0%	0%	0%	29%	29%
Tonic Water	**Alcampo**	21%	16%	2%	11%	1%	9%	9%	5%	13%	14%
	Eroski	0%	6%	0%	6%	18%	6%	0%	6%	52%	6%
	Mercadona	0%	45%	0%	21%	17%	17%	0%	0%	0%	0%
Sparkling Water	**Alcampo**	7%	15%	0%	20%	1%	5%	1%	7%	24%	20%
	Eroski	0%	0%	7%	0%	54%	0%	0%	0%	13%	26%
	Mercadona	0%	0%	0%	0%	43%	0%	0%	0%	29%	29%
Still Water	**Alcampo**	21%	16%	2%	11%	1%	9%	9%	5%	13%	14%
	Eroski	0%	6%	0%	6%	18%	6%	0%	6%	52%	6%
	Mercadona	0%	45%	0%	21%	17%	17%	0%	0%	0%	0%
Vegetable Milk	**Alcampo**	1%	0%	16%	19%	6%	0%	6%	12%	22%	18%
	Eroski	0%	26%	23%	6%	19%	0%	0%	6%	16%	4%
	Mercadona	0%	0%	0%	16%	0%	16%	0%	17%	34%	16%
Alcoholic Beer	**Alcampo**	8%	10%	8%	11%	7%	10%	4%	13%	22%	8%
	Eroski	20%	8%	6%	8%	17%	2%	6%	5%	14%	14%
	Mercadona	0%	23%	0%	15%	0%	31%	0%	8%	0%	23%
Overall	**Alcampo**	6%	10%	5%	13%	8%	7%	6%	11%	20%	13%
	Eroski	6%	8%	5%	6%	22%	5%	3%	7%	22%	15%
	Mercadona	0%	14%	0%	10%	12%	11%	1%	4%	8%	39%

Notes: Each cell represents the proportion of prices within a seller-category pair with a given digit in the second decimal position (last digit of the price). Different shades of yellow were used to highlight cells with a concentration of prices above the uniform distribution value (10%). In the bottom three rows, green colour was used to highlight the digits with a frequency above 10% in each supermarket pooling all categories together.

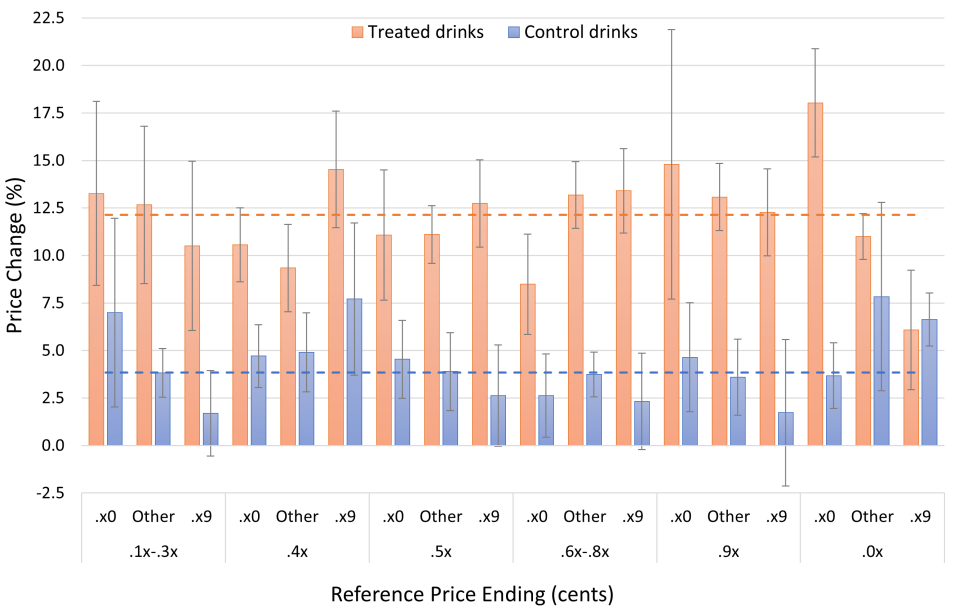
C.3 Additional figures

Figure C.1: Average price changes, by initial price ending (last digit)



Notes: This graph complements Figure 3.7 by illustrating the price changes of products with respect to their average price in December 2020 (before the tax change), by treatment group and initial (reference) price ending. The points in the plot were calculated through a linear combination of the estimates from column (3) in Table C.2, Appendix C.1.

Figure C.2: Average price changes, by initial price ending (two digits)



Notes: This graph complements Figure 3.9 by illustrating the price changes of products between December 2020 and February-April 2021, by treatment group and initial (reference) price ending. The values in this bar chart were calculated through a linear combination of the estimates from column (3) in Table C.4, Appendix C.1. The horizontal dashed lines denote the average change for each group of beverages (Treated and Control).

References

Bibliography

D. R. Agrawal and W. H. Hoyt. Pass-through in a multiproduct world. *Available at SSRN 3173180*, 2019.

A. Alesina and P. Giuliano. Preferences for redistribution. In *Handbook of social economics*, volume 1, pages 93–131. Elsevier, 2011.

H. Allcott, B. B. Lockwood, and D. Taubinsky. Should we tax sugar-sweetened beverages? an overview of theory and evidence. *Journal of Economic Perspectives*, 33(3):202–27, August 2019a.

H. Allcott, B. B. Lockwood, and D. Taubinsky. Regressive sin taxes, with an application to the optimal soda tax. *The Quarterly Journal of Economics*, 134 (3):1557–1626, 2019b.

E. T. Anderson and D. I. Simester. Effects of 9 price endings on retail sales: Evidence from field experiments. *Quantitative marketing and Economics*, 1 (1):93–110, 2003.

N. H. Anderson. A simple model for information integration. *Theories of cognitive consistency: A sourcebook*, pages 731–743, 1968.

T. Andreyeva, M. W. Long, and K. D. Brownell. The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food. *American journal of public health*, 100(2):216–222, 2010.

I. Ater and O. Gerlitz. Round prices and price rigidity: Evidence from outlawing odd prices. *Journal of Economic Behavior & Organization*, 144:188–203, 2017.

L. Bandy, P. Scarborough, R. Harrington, M. Rayner, and S. Jebb. Reductions in sugar sales from soft drinks in the uk from 2015 to 2018. *BMC medicine*, 18 (1):20, 2020.

R. Benabou. Inequality and growth. *NBER macroeconomics annual*, 11:11–74, 1996.

R. Bénabou and E. A. Ok. Social mobility and the demand for redistribution: the poum hypothesis. *NBER Working Paper*, (w6795), 1998.

S. Benoit, T. Schaefers, and R. Heider. Understanding on-the-go consumption: Identifying and quantifying its determinants. *Journal of Retailing and Consumer Services*, 31:32–42, 2016.

N. Berardi, P. Sevestre, M. Tepaut, and A. Vigneron. The impact of a 'soda tax'on prices: evidence from french micro data. *Applied Economics*, 48(41): 3976–3994, 2016.

G. Y. Bizer and R. M. Schindler. Direct evidence of ending-digit drop-off in price information processing. *Psychology & Marketing*, 22(10):771–783, 2005.

A. Blinder, E. R. Canetti, D. E. Lebow, and J. B. Rudd. *Asking about prices: a new approach to understanding price stickiness*. Russell Sage Foundation, 1998.

A. S. Blinder. Why are prices sticky? preliminary results from an interview study. Technical report, National Bureau of Economic Research, 1991.

B. Bollinger and S. Sexton. Local excise taxes, sticky prices, and spillovers: evidence from berkeley's soda tax. *Sticky Prices, and Spillovers: Evidence from Berkeley's Soda Tax (January 12, 2018)*, 2018.

C. Bonnet and V. Réquillart. Tax incidence with strategic firms in the soft drink market. *Journal of Public Economics*, 106:77–88, 2013.

C. J. Boyce, G. D. Brown, and S. C. Moore. Money and happiness: Rank of income, not income, affects life satisfaction. *Psychological science*, 21(4):471–475, 2010.

J. P. Bray and C. Harris. The effect of 9-ending prices on retail sales: A quantitative uk based field study. *Journal of Marketing Management*, 22(5-6):601–617, 2006.

A. D. Briggs, O. T. Mytton, A. Kehlbacher, R. Tiffin, M. Rayner, and P. Scarborough. Overall and income specific effect on prevalence of overweight and obesity of 20% sugar sweetened drink tax in uk: econometric and comparative risk assessment modelling study. *Bmj*, 347:f6189, 2013.

J. I. Bulow and P. Pfleiderer. A note on the effect of cost changes on prices. *Journal of political Economy*, 91(1):182–185, 1983.

C. C. Inquiry Group. Report on the anticipated acquisition by ag barr p.l.c. of britvic plc by means of an all-share merger. *The Competition Commission*, 2013.

R. M. Campos-Vázquez and E. M. Medina-Cortina. Pass-through and competition: the impact of soft drink taxes as seen through mexican supermarkets. *Latin American Economic Review*, 28(1):1–23, December 2019.

S. Capacci, O. Allais, C. Bonnet, and M. Mazzocchi. The impact of the french soda tax on prices and purchases. an ex post evaluation. *PLOS ONE*, 14(10): 1–22, 10 2019.

J. C. Caro, C. Corvalán, M. Reyes, A. Silva, B. Popkin, and L. S. Taillie. Chile's 2014 sugar-sweetened beverage tax and changes in prices and purchases of sugar-sweetened beverages: An observational study in an urban environment. *PLoS medicine*, 15(7):e1002597, 2018.

J. V. Castelló and G. L. Casasnovas. Impact of ssb taxes on sales. *Economics & Human Biology*, 36:100821, 2020.

J. Cawley and D. E. Frisvold. The pass-through of taxes on sugar-sweetened beverages to retail prices: The case of berkeley, california. *Journal of Policy Analysis and Management*, 36(2):303–326, 2017.

J. Cawley, B. Willage, and D. Frisvold. Pass-through of a tax on sugar-sweetened beverages at the philadelphia international airport. *Jama*, 319(3):305–306, 2018.

J. Cawley, D. Frisvold, A. Hill, and D. Jones. The impact of the philadelphia beverage tax on prices and product availability. *Journal of Policy Analysis and Management*, 39(3):605–628, 2020.

F. J. Chaloupka and H. Wechsler. Price, tobacco control policies and smoking among young adults. *Journal of Health Economics*, 16(3):359 – 373, 1997.

D. L. Chen, M. Schonger, and C. Wickens. {oTree} — {An} open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016.

C. W. Chow. The Effects of Job Standard Tightness and Compensation Scheme on Performance: An Exploration of Linkages. *The Accounting Review*, 58(4): 667–685, 1983.

J. Chu, M. Arce-Urriza, J.-J. Cebollada-Calvo, and P. K. Chintagunta. An empirical analysis of shopping behavior across online and offline channels for grocery products: The moderating effects of household and product characteristics. *Journal of Interactive Marketing*, 24(4):251 – 268, 2010.

M. A. Colchero, J. C. Salgado, M. Unar-Munguía, M. Molina, S. Ng, and J. A. Rivera-Dommarco. Changes in prices after an excise tax to sweetened sugar beverages was implemented in mexico: Evidence from urban areas. *PLOS ONE*, 10(12):1–11, 12 2015.

G. Cruces, R. Perez-Truglia, and M. Tetaz. Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics*, 98:100–112, 2013.

C. Cuadrado, J. Dunstan, N. Silva-Illanes, A. J. Mirelman, R. Nakamura, and M. Suhrcke. Effects of a sugar-sweetened beverage tax on prices and affordability of soft drinks in chile: A time series analysis. *Social Science & Medicine*, 245:112708, 2020. ISSN 0277-9536.

P. J. Danaher, I. W. Wilson, and R. A. Davis. A comparison of online and offline consumer brand loyalty. *Marketing Science*, 22(4):461–476, 2003.

A. M. Degeratu, A. Rangaswamy, and J. Wu. Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of research in Marketing*, 17(1): 55–78, 2000.

S. Delipalla and M. Keen. The comparison between ad valorem and specific taxation under imperfect competition. *Journal of Public Economics*, 49(3): 351–367, 1992.

S. Dharmasena, G. C. Davis, and O. Capps Jr. Partial versus general equilibrium calorie and revenue effects associated with a sugar-sweetened beverage tax. *Journal of Agricultural and Resource Economics*, pages 157–173, 2014.

S. A. Drakopoulos. Psychological thresholds, demand and price rigidity. *The Manchester School*, 60(2):152–168, 1992.

P. Dubois, R. Griffith, and M. O'Connell. How well targeted are soda taxes? *American Economic Review*, 110(11):3661–3704, 2020.

W. H. Dutton, G. Blank, and D. Groselj. *Cultures of the internet: the internet in Britain: Oxford Internet Survey 2013 Report*. Oxford Internet Institute, 2013.

F. Etilé, S. Lecocq, and C. Boizot-Szantai. The incidence of soft-drink taxes on consumer prices and welfare: Evidence from the french "soda tax". Pse working papers, HAL, 2018.

M. D. Evans and J. Kelley. Subjective social location: Data from 21 nations. *International Journal of Public Opinion Research*, 16(1):3–38, 2004.

J. Falbe, N. Rojas, A. H. Grummon, and K. A. Madsen. Higher retail prices of sugar-sweetened beverages 3 months after implementation of an excise tax in berkeley, california. *American journal of public health*, 105(11):2194–2201, 2015.

J. Fernández-Albertos and A. Kuo. Income perception, information, and progressive taxation: Evidence from a survey experiment. *Political Science Research and Methods*, 6(1):83–110, 2018.

G. Franz. Price effects on the smoking behaviour of adult age groups. *Public Health*, 122(12):1343 – 1348, 2008.

V. Gimpelson and D. Treisman. Misperceiving inequality. *Economics & Politics*, 30(1):27–54, 2018.

U. Gneezy, M. Niederle, and A. Rustichini. Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3): 1049–1074, 2003.

J. Gonçalves and J. P. Dos Santos. Brown sugar, how come you taste so good? the impact of a soda tax on prices and consumption. *Social Science & Medicine*, 264:113332, 2020. ISSN 0277-9536.

Y. Gorodnichenko and O. Talavera. Price setting in online markets: Basic facts, international comparisons, and cross-border integration. *American Economic Review*, 107(1):249–82, January 2017.

L. H. Goulder and R. C. Williams III. The substantial bias from ignoring general equilibrium effects in estimating excess burden, and a practical solution. *Journal of political Economy*, 111(4):898–927, 2003.

J. Grogger. Soda taxes and the prices of sodas and other drinks: Evidence from mexico. *American Journal of Agricultural Economics*, 99(2):481–498, 2017.

O. P. Hauser and M. I. Norton. (mis) perceptions of inequality. *Current opinion in psychology*, 18:21–25, 2017.

Health and Social Care Information Centre. Statistics on obesity, physical activity and diet, england, 2020. *Retrieved Government Statistics Service website. Accessed 30/10/2020*, 2020.

R. Herrmann, A. Moeser, and S. A. Weber. Price rigidity in the german grocery-retailing sector: scanner-data evidence on magnitude and causes. *Journal of Agricultural & Food Industrial Organization*, 3(1), 2005.

J. V. Hinrichs, D. S. Yurko, and J.-M. Hu. Two-digit number comparison: Use of place information. *Journal of Experimental Psychology: Human Perception and Performance*, 7(4):890, 1981.

G. H. Hong and N. Li. Market structure and cost pass-through in retail. *Review of Economics and Statistics*, 99(1):151–166, 2017.

R. M. Hower. *History of Macy's of New York, 1853-1919*. Harvard University Press, 1943.

K. B. Hvidberg, C. T. Kreiner, and S. Stantcheva. Social position and fairness views. *NBER Working Paper*, (w28099), 2020.

F. Jenkin. 3. on the principles which regulate the incidence of taxes. *Proceedings of the Royal Society of Edinburgh*, 7:618–631, 1872.

D. Kahneman and A. Tversky. Subjective probability: A judgment of representativeness. *Cognitive psychology*, 3(3):430–454, 1972.

M. Karadja, J. Mollerstrom, and D. Seim. Richer (and holier) than thou? the effect of relative income improvements on demand for redistribution. *Review of Economics and Statistics*, 99(2):201–212, 2017.

L. Kenworthy and L. McCall. Inequality, public opinion and redistribution. *Socio-Economic Review*, 6(1):35–68, 2007.

E. S. Knotek. Convenient prices and price rigidity: cross-sectional evidence. *Review of Economics and Statistics*, 93(3):1076–1086, 2011.

E. S. Knotek, D. Sayag, and A. Snir. The effects of price endings on price rigidity: Evidence from vat changes. 2019.

L. J. Kotlikoff and L. H. Summers. Tax incidence. In *Handbook of public economics*, volume 2, pages 1043–1092. Elsevier, 1987.

L. M. Kreul. Magic numbers: Psychological aspects of menu pricing. *Cornell Hotel and Restaurant Administration Quarterly*, 23(2):70–75, 1982.

I. Kuziemko, M. I. Norton, E. Saez, and S. Stantcheva. How elastic are preferences for redistribution? Evidence from randomized survey experiments. *The American Economic Review*, 105(4):1478–1508, 2015.

A. P. Lerner. Economic theory and socialist economy. *The Review of Economic Studies*, 2(1):51–61, 1934.

D. Levy, D. Lee, H. Chen, R. J. Kauffman, and M. Bergen. Price points and price rigidity. *Review of Economics and Statistics*, 93(4):1417–1431, 2011.

M. Lopez-Pastor, J. Garcia-Madariaga, J. Sánchez, and J. Figueiredo. Demand impact for prices ending with "9" and "0" in online and offline consumer goods retail trade channels. *International Review of Management and Marketing*, 10 (6):58–78, 2020.

A. H. Meltzer and S. F. Richard. A rational theory of the size of government. *Journal of political Economy*, 89(5):914–927, 1981.

K. L. Milkman, T. Rogers, and M. H. Bazerman. I'll have the ice cream soon and the vegetables later: A study of online grocery purchases and order lead time. *Marketing Letters*, 21(1):17–35, 2010.

K. B. Monroe. Measuring price thresholds by psychophysics and latitudes of acceptance. *Journal of marketing research*, 8(4):460–464, 1971.

K. B. Monroe. Buyers' subjective perceptions of price. *Journal of marketing research*, 10(1):70–80, 1973.

K. B. Monroe. Pricing: Making profitable decisions. 2003.

A. Muhammad, B. Meade, D. R. Marquardt, and D. Mozaffarian. Global patterns in price elasticities of sugar-sweetened beverage intake and potential effectiveness of tax policy: a cross-sectional study of 164 countries by sex, age and global-income decile. *BMJ Open*, 9(8), 2019.

M. O'Connell and K. Smith. Corrective tax design and market power. 2020.

Office for National Statistics. Internet access – households and individuals, Great Britain, 2019. *https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2019*, Last accessed on 2020-11-06.

K. Pauwels, S. Srinivasan, and P. H. Franses. When do price thresholds matter in retail categories? *Marketing Science*, 26(1):83–100, 2007.

A. C. Pigou. *The Economics of Welfare*. McMillan & Co., London, 1920.

T. Piketty. Social mobility and redistributive politics. *The Quarterly journal of economics*, 110(3):551–584, 1995.

S. E. Poltrock and D. R. Schwartz. Comparative judgments of multidigit numbers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1):32, 1984.

L. M. Powell, J. F. Chriqui, T. Khan, R. Wada, and F. J. Chaloupka. Assessing the potential effectiveness of food and beverage taxes and subsidies for improving public health: a systematic review of prices, demand and body weight outcomes. *Obesity reviews*, 14(2):110–128, 2013.

L. M. Powell, J. Leider, and P. T. Léger. The impact of the cook county, il, sweetened beverage tax on beverage prices. *Economics & Human Biology*, 37: 100855, 2020.

Public Health England. Health matters: obesity and the food environment, 2017.

M. Rafiq and H. Fulford. Loyalty transfer from offline to online stores in the uk grocery industry. *International Journal of Retail & Distribution Management*, 2005.

C. Rojas and E. Y. Wang. Do taxes for soda and sugary drinks work? scanner data evidence from berkeley and washington. *Scanner Data Evidence from Berkeley and Washington (September 23, 2017)*, 2017.

P. Scarborough, V. Adhikari, R. A. Harrington, A. Elhussein, A. Briggs, M. Rayner, J. Adams, S. Cummins, T. Penney, and M. White. Impact of the announcement and implementation of the uk soft drinks industry levy on sugar content, price, product size and number of available soft drinks in the uk, 2015-19: A controlled interrupted time series analysis. *PLOS Medicine*, 17 (2):1–19, 02 2020.

R. M. Schindler. The 99 price ending as a signal of a low-price appeal. *Journal of Retailing*, 82(1):71–77, 2006.

R. M. Schindler and A. R. Wiman. Effects of odd pricing on price recall. *Journal of Business Research*, 19(3):165–177, 1989.

S. Seiler, A. Tuchman, and S. Yao. The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects. *Journal of Marketing Research*, 58(1): 22–49, 2021.

V. Shankar, A. K. Smith, and A. Rangaswamy. Customer satisfaction and loyalty in online and offline environments. *International Journal of Research in Marketing*, 20(2):153 – 175, 2003. ISSN 0167-8116. doi: https://doi.org/10.1016/S0167-8116(03)00016-8. URL *http://www.sciencedirect.com/science/article/pii/S0167811603000168*.

L. D. Silver, S. W. Ng, S. Ryan-Ibarra, L. S. Taillie, M. Induni, D. R. Miles, J. M. Poti, and B. M. Popkin. Changes in prices, sales, consumer spending, and beverage consumption one year after a tax on sugar-sweetened beverages in berkeley, california, us: A before-and-after study. *PLoS medicine*, 14(4): e1002283, 2017.

G. J. Stigler. The economics of information. *Journal of political economy*, 69(3): 213–225, 1961.

A. Strulov-Shlain. More than a penny's worth: Left-digit bias and firm pricing. *Chicago Booth Research Paper*, (19-22), 2019.

M. Sutter and H. Weck-Hannemann. Taxation and the veil of ignorance - A real effort experiment on the laffer curve. *Public Choice*, 115(1-2):217–240, 2003.

M. Thomas and V. Morwitz. Penny wise and pound foolish: the left-digit effect in price cognition. *Journal of Consumer Research*, 32(1):54–64, 2005.

R. Tiffin, A. Kehlbacher, and M. Salois. The effects of a soft drink tax in the uk. *Health Economics*, 24(5):583–600, 2015.

D. W. Twedt. Does the" 9 fixation" in retail pricing really promote sales? *Journal of Marketing (pre-1986)*, 29(000004):54, 1965.

G. van Rossum. Python tutorial. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.

E. G. Weyl and M. Fabinger. Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy*, 121(3): 528–583, 2013.

WHO. *Using price policies to promote healthier diets*. WHO Regional Office for Europe, 2015a.

WHO. Guideline: sugars intake for adults and children. *World Health Organization*, 2015b. URL *https://apps.who.int/iris/bitstream/handle/10665/149782/9789241549028_eng.pdf*.

WHO. Fiscal policies for diet and prevention of noncommunicable diseases: technical meeting report. *World Health Organisation*, 2016. URL *http://apps.who.int/iris/bitstream/10665/250131/1/9789241511247-eng.pdf*.